PHYLOGENETIC INFERENCE IN SAXIFRAGACEAE SENSU STRICTO AND *GILIA* (POLEMONIACEAE) USING *mat*K SEQUENCES[1]

Leigh A. Johnson and Douglas E. Soltis[2]

ABSTRACT

Comparative sequencing of the maturase-encoding chloroplast gene *mat*K has great potential for reconstructing phylogenetic relationships not only within families, but also within genera of land plants. This gene of 1550 bp is easily amplified due to highly conserved, flanking coding regions that include the *trn*K exons, *rps*16, and *psb*A. Several available sequencing primers also have wide applicability. Parsimony analysis of 45 *mat*K sequences representing Saxifragaceae sensu stricto provides a level of resolution comparable to that obtained via chloroplast DNA restriction site analysis. Furthermore, this analysis suggests relationships among genera and species that are highly concordant with the results of separate analyses of *rbc*L sequences and chloroplast DNA restriction sites, and with those of combined analyses of these three chloroplast DNA data sets. Parsimony analysis of 31 *mat*K sequences representing all six sections of *Gilia* (Polemoniaceae) and 10 allied genera provides strong evidence for the polyphyly of *Gilia* and suggests relationships among sections of *Gilia* that are highly concordant with a recent ITS sequence analysis of the Polemoniaceae. Our analyses suggest that *mat*K sequences are not strongly biased toward transitions, and the frequency of mutations at the first and second codon positions approach the frequency of mutations in the third codon position.

Investigation of the chloroplast genome, either through analysis of restriction site mutations, structural rearrangements, or DNA sequences, has dominated plant molecular systematic research during the past decade. These approaches have proven extremely useful in addressing a broad range of systematic and evolutionary questions at all levels of taxonomic hierarchy. Of these approaches, comparative sequencing of chloroplast, as well as nuclear, genes has become particularly popular in recent years, due in large part to the relative ease of generating sequences and the unambiguity of the data. The large number of recent systematic studies employing sequencing of the chloroplast gene *rbc*L attests to the enormous phylogenetic potential of comparative sequencing (e.g., Brunsfeld et al., 1994; Chase et al., 1993; Conti et al., 1993; Donoghue et al., 1992; Gadek & Quinn, 1993; Giannasi et al., 1992; Kim et al., 1992; Kron & Chase, 1993; Olmstead et al., 1993; Morgan & Soltis, 1993; Price & Palmer, 1993; Qiu et al., 1993; Rodman et al., 1993; Smith et al., 1993; Soltis et al., 1990). Although *rbc*L sequence analysis has overshadowed the use of other gene sequences in plant systematics, the phylogenetic utility of several other DNA regions, both nuclear and organellar, has been investigated in plants. Among chloroplast genes, *atpβ* (Ritland & Clegg, 1987; Hoot et al., 1995 this issue), *mat*K (Johnson & Soltis, 1994; Steele & Vilgalys, 1994), and *ndh*F (Olmstead & Sweere, 1994; Olmstead & Reeves, 1995, this issue) provide regions with demonstrated utility for inferring phylogenies at taxonomic levels unresolvable with *rbc*L sequences alone. The use of these, and other regions of chloroplast DNA (cpDNA), such as noncoding intergenic spacers (Gielly & Taberlet, 1994; Golenberg et al., 1993), for phylogenetic inference suggests that analysis of the chloroplast genome will continue to provide important information for systematics.

Among protein coding regions in the chloroplast genome, *mat*K (ORFK) is one of the most rapidly evolving (Wolfe, 1991). The chloroplast gene *mat*K

[1] We thank Paul Gadek, Christopher Quinn, Clifford Morden, and Jeff Palmer for sharing unpublished results of their studies involving *mat*K; John Wiens and Tod Reeder for providing their unpublished manuscript on "pseudofossils"; Mark Porter for providing information and discussion concerning his ITS sequence analysis of the Polemoniaceae; Robert Jansen for providing *ndh*F sequences for comparative purposes; Mark Chase, Paul Gadek, and Pam Soltis for providing helpful comments on an earlier draft of this manuscript; and the many individuals who provided plant material used for sequencing. This research was supported by NSF grants BSR 9007614 and DEB 9307000 to DES, DEB 9321788 to LAJ and DES, and grants from Sigma Xi, the American Society of Plant Taxonomists, and the California Native Plant Society to LAJ.

[2] Department of Botany, Washington State University, Pullman, Washington 99164-4238, U.S.A.

is approximately 1550 base pairs (bp) in length and encodes a maturase involved in splicing type II introns from RNA transcripts (Neuhaus & Link, 1987; Wolfe et al., 1992). In all photosynthetic land plants examined to date, *mat*K is located within an intron of approximately 2600 bp located between the 5′ and 3′ exons of the transfer RNA gene for lysine, *trn*K (Fig. 1). However, in the reduced chloroplast genome of the nonphotosynthetic parasite *Epifagus virginiana* (L.) Barton (Orobanchaceae), neither the *trn*K coding regions nor associated intron are present (Wolfe et al., 1992). In *Epifagus*, *mat*K is bounded by *trn*Q and the *psb*A pseudogene. The presence of *mat*K in the very reduced plastid genome of *Epifagus* suggests that *mat*K is functionally important and has a broader intron splicing role than simply splicing the *trn*K intron in which it typically occurs (Wolfe et al., 1992). Alignment of seven complete *mat*K amino acid sequences representing bryophytes, gymnosperms, monocots, and dicots reveals the presence of more conserved regions interrupted by stretches of sequence having little similarity (C. Morden and J. Palmer, pers. comm.).

We recently reported that *mat*K sequences of only 750 bp provided resolution of relationships in Saxifragaceae sensu stricto (s. s.) comparable to that obtained via restriction site analysis of the entire chloroplast genome based on 20 endonucleases; these *mat*K sequences provided greater resolution than did *rbc*L sequences (Johnson & Soltis, 1994). Herein we explore further the phylogenetic utility of *mat*K sequence variation through: (1) phylogenetic analyses of an expanded data set for Saxifragaceae s. s., a preliminary data set for *Gilia* (Polemoniaceae), and analysis of the polyploid origin of *Saxifraga osloensis* Knaben; (2) description of *mat*K sequence variation; and (3) discussion of the broad applicability of *mat*K sequencing primers in various plant groups.

Saxifragaceae s. s. continue to provide an ideal opportunity for assessing the utility of other chloroplast data sets because the family is well defined by molecular data, and both *rbc*L sequences and cpDNA restriction sites have been previously gathered for virtually all genera (e.g., Soltis et al., 1991; Soltis et al., 1993; Morgan & Soltis, 1993). We have expanded our *mat*K sequence matrix for Saxifragaceae s. s. from that reported by Johnson & Soltis (1994) to include an additional 324 bp and 14 additional species yielding 1078 bp of sequence data for 45 taxa. Additional species of *Saxifraga* and *Chrysosplenium* have been included because these genera possessed the longest branch lengths in our previous study (Johnson & Soltis, 1994). Additional members of the *Heuchera* and *Boykinia* groups of genera were also included to explore more fully the degree of resolution obtainable with *mat*K sequences in these well-supported groups. We present analyses of *mat*K data alone, as well as the results of combined analyses involving *mat*K and *rbc*L sequences and cpDNA restriction sites (Soltis et al., 1993) to provide a comprehensive view of relationships in Saxifragaceae s. s. as suggested by these three cpDNA data sets. We also discuss the fine-scale resolution that *mat*K can provide using an example of polyploid evolution in *Saxifraga*.

In addition to Saxifragaceae s. s., we also illustrate the phylogenetic utility of *mat*K sequences using analyses of a preliminary data set for *Gilia* (Polemoniaceae). Steele & Vilgalys (1994) presented an analysis using *mat*K sequences to resolve relationships among genera of Polemoniaceae, but they did not address intrageneric relationships within any of these genera. *Gilia* is a large, morphologically diverse genus of questionable monophyly and affinities. In many respects, the approximately 70 species currently recognized in *Gilia* (Day, 1993a) are united more by the lack of synapomorphic characters that circumscribe any of the other 12 to 14 temperate genera rather than by any consistent set of characters unique to themselves. We compare our phylogenetic trees for *Gilia* based on *mat*K sequences to those obtained in a recent analysis of sequences obtained from the nuclear ribosomal DNA internal transcribed spacer regions (rDNA-ITS) that included this genus (Porter, 1993).

We describe the nature of sequence variation in *mat*K, including: (1) comparison of nucleotide variability between *mat*K and *rbc*L, *ndh*F, and the ITS regions; (2) comparison of transition : transversion ratios and substitution rates by codon position between *mat*K and *rbc*L; (3) determination of the degree of random structuring of variation (Archie, 1989a); and (4) determination of the number and phylogenetic distribution of insertion-deletion events.

Lastly, in an effort to promote the use of *mat*K by other investigators, we also discuss the applicability of PCR and sequencing primers in various taxonomic groups.

Materials and Methods

Amplification and Sequencing

DNA was isolated from all taxa (Appendix 1) using a CTAB buffer method (Doyle & Doyle, 1987) as modified by Soltis et al. (1991). Ampli-

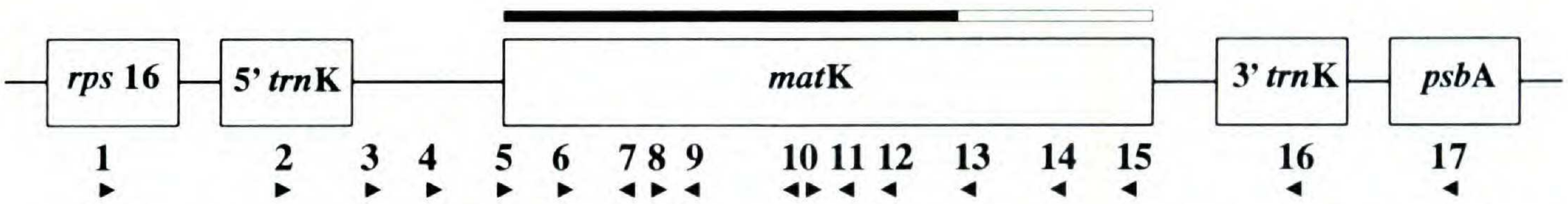


FIGURE 1. Relative location of amplification and sequencing primers used in this study for sequencing *mat*K and the *trn*K intron. Numbers refer to primers listed in Table 1. The shaded region above *mat*K indicates the region sequenced for species included in the Saxifragaceae s. s. and *Gilia* data matrices.

fication of *mat*K was accomplished via the polymerase chain reaction (PCR) to obtain sufficient quantities of DNA for sequencing. Several synthetic PCR primers located in the 5′*trn*K and 3′*trn*K exons, as well as in the *rps*16 and *psb*A genes that flank *trn*K (and thus *mat*K; Fig. 1, Table 1), have been designed. Two PCR primer pairs (*rps*16-4547F and *trn*K-2R; and *trn*K-3914F and *psb*A-R) provide DNA amplification products for initial sequencing of the entire *trn*K intron using *trn*K-3914F and *trn*K-2R, respectively, as sequencing primers (e.g., Johnson & Soltis, 1994). In this fashion, we sequenced the entire *trn*K intron in *Bensoniella oregona*, *Saxifraga integrifolia*, and *Sullivantia sullivantii* (a complete *mat*K sequence, excluding the flanking *trn*K intron regions, was also generated for *Sullivantia oregana*). For all other taxa, we used *trn*K-3914F and *trn*K-2R to produce double-stranded DNA from total genomic DNA. We subsequently used this double-stranded DNA as a template to produce single-stranded DNAs using primer *trn*K-3914F individually for forward strand synthesis, and primer *trn*K-2R for reverse strand synthesis. All PCR reactions used Replitherm polymerase from Epicentre Technologies following the manufacturer's suggested concentrations of all reagents and DNA. The PCR temperature profile we employed consisted of 30 cycles at 94°C for 1 minute 30 seconds, 48°C for 2 minutes, and 72°C for 3 minutes with an additional 15 minutes at 72°C following the final cycle. Secondary bands were occasionally observed in agarose test gels of our single-stranded DNA amplification products but only rarely in our double-stranded DNA products. These secondary bands never posed any problems during sequencing.

TABLE 1. Base composition of amplification and sequencing primers discussed in this study. Primers *rps*16-4547F and *trn*K-3914F (monocot) were designed by Jerry Learn. We later modified *trn*K-3914F (monocot) to produce *trn*K-3914F (dicot) and hence provide greater homology to dicots. Primers *trn*K-2R and *psb*A-R were designed by Kelly Steele. All other primers were designed in our lab and named based on their approximate position on the *trn*K map for *Sinapis* (Neuhaus & Link, 1987). Primer reference numbers correspond to those used in Figure 1.

Reference number in Figure 1	Name	5′ Sequence 3′
1	*rps*16-4547F	AGG TGC TCA ACC TAC AAG AAC C
2	*trn*K-3914F (dicot)	GGG GTT GCT AAC TCA ACG G
2	*trn*K-3914F (monocot)	ATC TGG GTT GCT AAC TCA ATG G
3	*trn*K-253F	TTG GGT CGA GTC AAT AAA T
4	*trn*K-582F	CTA ACC ATC TTG TTA TCC T
5	*trn*K-710F	GTA TCG CAC TAT GT[T/A] TCA TTT GA
6	*mat*K-934F	ATT TTG GTT ATG ACA ATA A
7	*mat*K-1168R	ATT GAA TGA ATT GAT CGT A
8	*mat*K-1176F	CAA TTC ATT CA[A/C] TAT TTC CTT
9	*mat*K-1235R	GG[A/G] GTG GGG TAT TAG TAT A
10	*mat*K-1412F	ATA TAA TTC TTA TGT ATG TG
10	*mat*K-1412R	CAC ATA [G/C]AT AAG AAT TAT AT
11	*mat*K-1470R	AAG ATG TTG AT[T/C] GTA AAT GA
12	*mat*K-1506R	TTC CAT AGA AAT ATA TTC G
13	*mat*K-1848R	TAT CGA ACT TCT TAA TAG C
14	*mat*K-2000R	ATT TCT GCA TAT GCG CAC AAA TC
15	*mat*K-2200R	TCT GTA TAA CCT CCA CAA AG
16	*trn*K-2R	AAC TAG TCG GAT GGA GTA G
17	*psb*A-R	CGC GTC TCT CTA AAA TTG CAG TCA T

Following PCR amplification, single-stranded DNAs were precipitated with 20% PEG/2.5M NaCl, washed in 70% and 95% EtOH, dried, and resuspended in TE (Morgan & Soltis, 1993). We subsequently employed dideoxy sequencing of the resuspended PCR products using the Sequenase 2.0 kit (U. S. Biochemical Corp.) and a set of sequencing primers (primers *mat*K-1168R or *mat*K-1235R, *mat*K-1470R, and *mat*K-1412F; or *mat*K-1168R, *mat*K-1506R, and *mat*K-1848R; Fig. 1; Table 1). These primers enable us to sequence routinely over two-thirds of *mat*K, beginning at the 5′ end (Fig. 1). Approximately 500 bp at the 3′ end were not sequenced because we have obtained sufficient resolution to address our present inquiries with the approximately 1080 bp obtained using the sequencing primers described above.

PLANT SAMPLES

Saxifragaceae s. s. In addition to extending our previously reported *mat*K sequences (Johnson & Soltis, 1994) by 324 bp, we sequenced 14 new species for a total of 45 sequences representing virtually all of Saxifragaceae s. s. (Appendix 1). Despite repeated attempts, we have been unable to obtain suitable material for DNA isolation from the monotypic genera *Hieronymusia*, *Saxifragella*, and *Zahlbrucknera* due to their geographically remote and restricted distributions. *Oresitrophe* is also missing from this analysis because we were not able to obtain material from this monotypic genus until after the lengthy phylogenetic analyses of Saxifragaceae s. s. were completed. *Tetracarpaea* and *Ribes* (Grossulariaceae) were included as outgroups because previous studies have shown these genera to be close relatives of Saxifragaceae s. s. (Morgan & Soltis, 1993; Soltis et al., 1993). Partial sequences of *Saxifraga osloensis*, *S. tridactylites*, and *S. adscendens* have also been obtained and are discussed below as an example of the insights *mat*K sequences can provide regarding polyploid origins; these sequences were not, however, included in our broad parsimony analyses of Saxifragaceae s. s.

Partial *mat*K sequences for Saxifragaceae s. s. vary from 1039 to 1063 bp in length and provide a matrix of 1078 characters after alignment. Ten insertion-deletion events (indels) of three, six, or nine nucleotides distributed among 18 species account for the length variation in these sequences (Appendix 2). We easily aligned these sequences visually and positioned indels so as to minimize base substitutions while maintaining the proper reading of codons. After alignment, we checked the reading frame and position of indels by translating the sequences to amino acids using either MacClade version 3.01 (Maddison & Maddison, 1992) or MEGA version 1.01 (Kumar et al., 1993). We scored missing bases associated with indels as ambiguous ("?"), rather than as a fifth character-state, and considered the phylogenetic distribution of each indel a posteriori by mapping its occurrence on trees derived from analysis of base substitutions alone.

These *mat*K sequences are the third chloroplast DNA data set constructed for the purpose of resolving relationships within Saxifragaceae s. s. A combined analysis of the other two data sets, cpDNA restriction sites and *rbc*L sequences, has recently been reported (Soltis et al., 1993). Here we compare and also combine our *mat*K sequences with these two other DNA data sets to obtain a comprehensive view of relationships suggested by cpDNA data in Saxifragaceae s. s. Because slight differences exist in the taxa sampled for the three molecular analyses (cpDNA restriction sites and *mat*K and *rbc*L sequences), we constructed three different combined matrices (Appendix 1) in an effort to obtain the most comprehensive combined analysis possible. Matrix-1 is the "purest" combined data set comprising 21 species for which all three character sets exist. Matrix-2 comprises all of the taxa from matrix-1 with the addition of *Astilbe* and *Chrysosplenium*, genera for which different species were analyzed in the various DNA studies. A complete data set for *Astilbe* was formed by combining the *rbc*L sequence and restriction site data from *A. taquetii* with the *mat*K sequence from *A. japonica* × *chinesensis*. Similarly, a complete data set for *Chrysosplenium* was formed by combining *mat*K and *rbc*L sequences from *C. iowense* with restriction site data from *C. americanum*. This approach seemed reasonable given that both *Astilbe* and *Chrysosplenium* are distinctive, well-defined genera in Saxifragaceae. We thus feel the likelihood of conflict among the character sets for these genera is not substantially greater than that existing among character sets for any of the other taxa for which a single species was used to generate all three data sets. Matrix-3 contains the broadest sampling of saxifragaceous genera, including all of the taxa in matrix-1 and an additional 22 taxa for which any two of the three character sets were available. For example, *Suksdorfia violacea* is included in matrix-3 because both cpDNA restriction site and *mat*K sequence data are available but not *rbc*L sequence data. Similarly, *Astilbe* and *Chrysosplenium* are represented in this matrix only by the character sets obtained from *A. taquetii* (cpDNA

restriction site and *rbc*L sequence data) and *C. iowense* (*mat*K and *rbc*L sequence data), rather than by the composite data set used in matrix-2. We filled the missing character set for each of the 11 taxa for which a character set is missing with "?" to indicate ambiguity for all missing characters.

Gilia. *Gilia* comprises approximately 70 species in six sections (sensu Day, 1993a, b). Although we are interested primarily in intersectional relationships involving sections *Saltugilia* and *Kellogia*, *Gilia* has served as the catch-all among temperate Polemoniaceae (Grant, 1959; Wherry, 1940), and the *mat*K sequence results (see below) underscore the necessity of broad sampling among allied genera to gain the clearest picture of relationships in *Gilia*. Our preliminary data matrix for *Gilia* includes 31 species representing all six sections of *Gilia* (sensu Day, 1993a, b) and eight allied temperate genera (*Allophyllum*, *Collomia*, *Eriastrum*, *Ipomopsis*, *Langloisia*, *Navarretia*, *Phlox*, and *Polemonium*; Appendix 1). We selected the tropical genera *Bonplandia* and *Cantua* as outgroups because both traditional (e.g., Grant, 1959) and molecular data (Porter, 1993; Steele & Vilgalys, 1994) suggest these genera are close allies of the temperate Polemoniaceae.

The *mat*K sequences generated for *Gilia* vary from 1065 to 1080 bp in length and provide a matrix of 1083 characters after alignment. Four indels of three, six, or nine nucleotides in length distributed among eight taxa account for the length variation in these sequences (Appendix 2). As with the Saxifragaceae, we scored all missing bases as ambiguous ("?") and considered the phylogenetic distribution of indels after parsimony analysis of the base substitutions alone.

PHYLOGENETIC ANALYSES

PAUP (version 3.1.1, Swofford, 1991) installed on a Macintosh Centris computer was used to search for most parsimonious trees, perform bootstrap analyses (Felsenstein, 1985; see also Felsenstein & Kishino, 1993; Hillis & Bull, 1993) and decay analyses (Bremer, 1988; Donoghue et al., 1992), and to calculate the retention index (RI, Farris 1989a, b; see also Archie, 1989b, 1990) and consistency index (CI, Kluge & Farris, 1969) with autapomorphies retained. Parsimony analyses employed heuristic searches using TBR branch swapping, MULPARS, HOLD = 5 and SIMPLE addition. Five hundred replications of RANDOM addition with TBR branch swapping and MULPARS were also employed to search for multiple islands of most parsimonious trees (Maddison, 1991). Bootstrap analyses for each data set consisted of 100 replications using heuristic searches with TBR branch swapping, SIMPLE addition, and HOLD = 5. For our Saxifragaceae s. s. data set of *mat*K sequences, we additionally set a MAXTREE limit of 1000 trees per replicate after the computer ran out of memory on its 38th replication in an earlier analysis. Because few of the replications reached this limit and we used the bootstrap as only one estimate of support for monophyletic groups, we do not feel this restriction compromises the integrity of our results. Decay analyses were performed for all data sets using the following two approaches. For all data sets except the *mat*K sequence matrix for Saxifragaceae s. s. and the combined cpDNA data matrix-3, TBR branch swapping and 100 RANDOM addition heuristic searches that saved all trees up to five steps longer than the most parsimonious tree length were performed. Strict consensus trees were formed from these trees for each length greater than the most parsimonious tree length after filtering out trees of inappropriate length. This approach was inadequate, however, for the *mat*K sequence matrix and the combined cpDNA data matrix-3 for Saxifragaceae s. s. because of computer memory constraints and the large number of trees even a few steps greater than the most parsimonious tree length. For these data matrices, the decay analysis was conducted successively for each step greater than the most parsimonious tree length for up to five steps by saving only trees from heuristic searches that failed to satisfy a constraint topology imposed by the strict consensus of all trees found in the previous steps (Johnson & Soltis, 1994; Morgan et al., in press; Soltis & Kuzoff, in press).

VARIABILITY ANALYSES

Comparisons between genes. To determine the distribution of base substitutions and relative variability of regions within *mat*K, we used the IBM-PC program MEGA to make pairwise comparisons of nucleotide differences per site in complete *mat*K and *rbc*L sequences from three saxifragaceous taxa: *Bensoniella oregona*, *Saxifraga integrifolia*, and *Sullivantia oregana*. Pairwise comparisons of nucleotide differences and amino acid differences were also made among these three species and an additional 22 saxifragaceous taxa for which *rbc*L and over two-thirds of *mat*K have been sequenced. Separate parsimony analyses of these 25 *mat*K and *rbc*L DNA sequences were performed to estimate homoplasy as indicated by

CI and RI and to compare transition : transversion and codon position ratios as calculated by MacClade over a most parsimonious tree for each data set.

The increasing availability of DNA sequences from regions other than *rbc*L also enabled us to conduct two additional small-scale comparisons between *mat*K and other DNA sequences. In one such analysis, we made pairwise comparisons of nucleotide substitutions per site between *mat*K and *ndh*F for *Coriandrum* (Apiaceae), *Griselinia* (Cornaceae), and *Hedera* (Araliaceae) using *ndh*F sequences kindly provided by Robert Jansen. A similar comparison was made between *mat*K and ITS sequences (ITS-1 and ITS-2 combined) for *Gilia leptalea*, *G. scopulorum*, and *G. splendens*.

Variation in Saxifragaceae s. s. and Gilia matK matrices. We calculated the transition : transversion and codon position ratios as reconstructed over a most parsimonious tree for both the Saxifragaceae s. s. and *Gilia mat*K sequence matrices. The consistency indices for the Saxifragaceae s. s. and *Gilia mat*K data sets were also calculated and compared to expected CI values derived from the regression equation of Sanderson & Donoghue (1989) after first removing all autapomorphies from each data set with the aid of a spreadsheet program.

To assess how the observed variation in our Saxifragaceae s. s. and *Gilia mat*K matrices is structured, we employed the randomization test of Archie (1989a; see also Källersjö et al., 1992). The length of the most parsimonious trees from analyses of both data sets was compared to the distribution of most parsimonious trees derived from analyses of 100 data sets created by randomly permuting character states within characters via the SHUFFLE option in MacClade. Analyses of each randomized data set consisted of two heuristic searches employing first SIMPLE addition, and then CLOSEST addition, HOLD = 5, MULPARS, TBR branch swapping, and MAXTREES = 50. The shorter of the two results was used as the most parsimonious tree length. This approach provided sufficient speed for determination of non-random structure in our large data sets and was more likely (based on our initial tests) to obtain a shorter average tree length from the sample of 100 randomized data sets than an approach using thorough (i.e., no MAXTREE limit) searches employing only SIMPLE or only CLOSEST addition. The average length of the most parsimonious trees from these randomized data sets was also used to calculate the homoplasy excess ratio (HER; Archie, 1989b) for both Saxifragaceae s. s. and *Gilia*.

Results

Phylogenetic Analyses

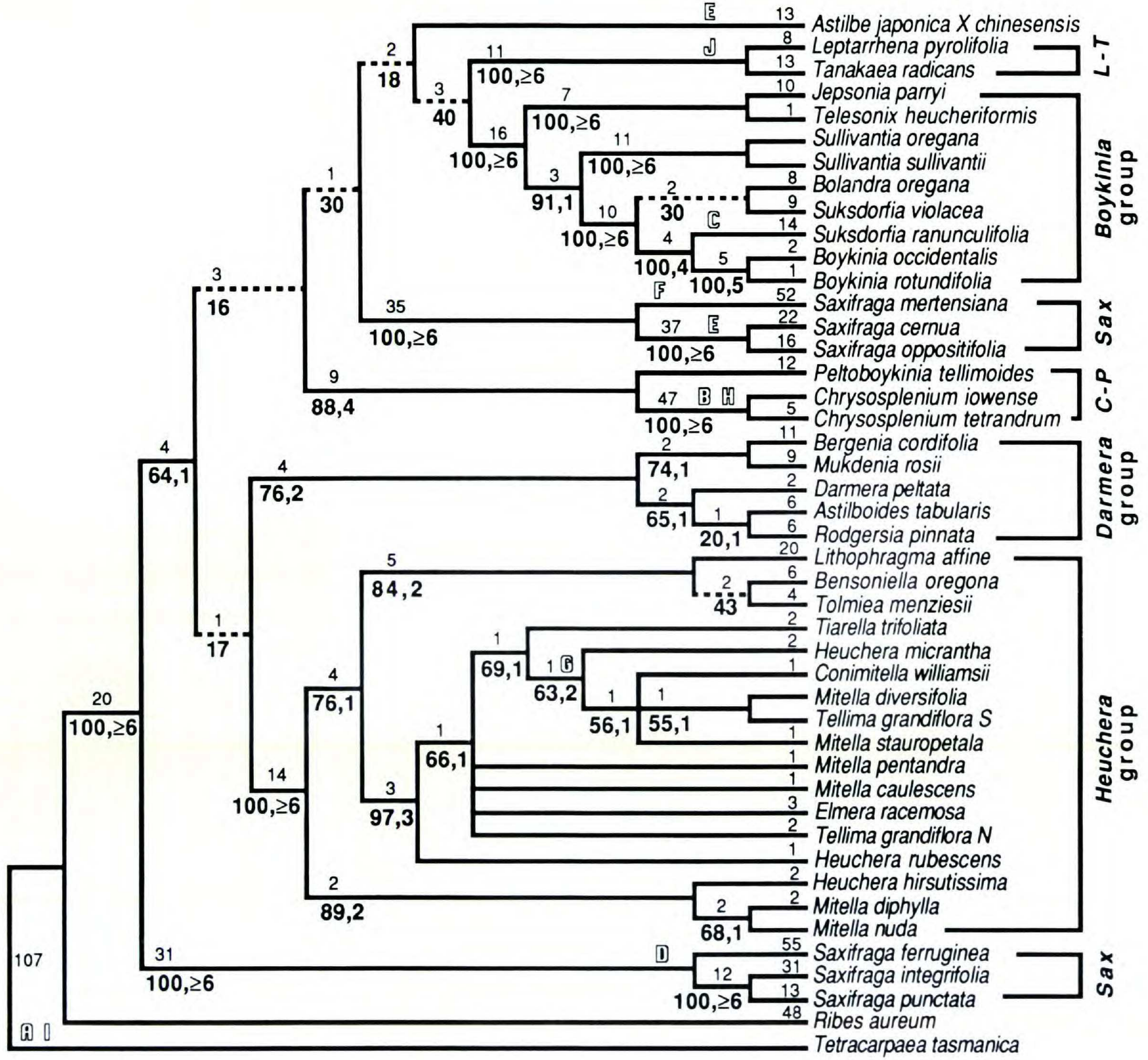
In analyses of all matrices, bootstrap and decay values generally agree in indicating support for monophyletic groups. Decay values ≥5 are rarely found on branches with bootstrap support of less than 95%, and bootstrap values ≥95% are rarely found on branches with decay values ≤3 (Figs. 2–6).

Saxifragaceae s. s. Parsimony analysis of *mat*K sequences for Saxifragaceae s. s. resulted in two islands with a total of 372 trees of 842 steps (Fig. 2). The consistency index excluding autapomorphies for these trees is 0.583, a higher value than that expected for 45 taxa using the regression equation of Sanderson & Donoghue (1989; Table 2). The large number of most parsimonious trees is primarily the result of homoplasy among the small number of substitutions supporting the basal branches. Increased homoplasy caused by the inclusion of a few highly divergent taxa also contributes to the large number of most parsimonious trees obtained. For example, although *Saxifraga mertensiana* is united with *S. cernua* and *S. oppositifolia* in all most parsimonious trees and the three taxa share 35 base substitutions (Fig. 2), removing *S. mertensiana* prior to parsimony analysis decreases the number of most parsimonious trees to 126.

Parsimony analyses of the combined cpDNA restriction site and *mat*K and *rbc*L sequence data sets for Saxifragaceae s. s. all yielded single islands of most parsimonious trees. Matrix-1 yielded two most parsimonious trees of 988 steps with a CI of 0.801 (autapomorphies retained; Fig. 3). Matrix-2 yielded a single most parsimonious tree of 1137 steps with a CI of 0.764 (autapomorphies retained;

→

FIGURE 2. One of 372 most parsimonious trees from analysis of *mat*K sequences for Saxifragaceae s. s. Base substitutions (ACCTRAN) are indicated above branches. Bootstrap and decay values are indicated below branches, respectively. Dashed lines represent branches that are not supported by all most parsimonious trees; in these instances the decay value is zero and is not indicated following the bootstrap value below the branch. Letters (A-J) denote the distribution of specific indels referenced in Appendix 2.


Astilbe japonica X chinesensis
Leptarrhena pyrolifolia
Tanakaea radicans
Jepsonia parryi
Telesonix heucheriformis
Sullivantia oregana
Sullivantia sullivantii
Bolandra oregana
Suksdorfia violacea
Suksdorfia ranunculifolia
Boykinia occidentalis
Boykinia rotundifolia
Saxifraga mertensiana
Saxifraga cernua
Saxifraga oppositifolia
Peltoboykinia tellimoides
Chrysosplenium iowense
Chrysosplenium tetrandrum
Bergenia cordifolia
Mukdenia rosii
Darmera peltata
Astilboides tabularis
Rodgersia pinnata
Lithophragma affine
Bensoniella oregona
Tolmiea menziesii
Tiarella trifoliata
Heuchera micrantha
Conimitella williamsii
Mitella diversifolia
Tellima grandiflora S
Mitella stauropetala
Mitella pentandra
Mitella caulescens
Elmera racemosa
Tellima grandiflora N
Heuchera rubescens
Heuchera hirsutissima
Mitella diphylla
Mitella nuda
Saxifraga ferruginea
Saxifraga integrifolia
Saxifraga punctata
Ribes aureum
Tetracarpaea tasmanica
L-T
Boykinia group
Sax
C-P
Darmera group
Heuchera group
Sax

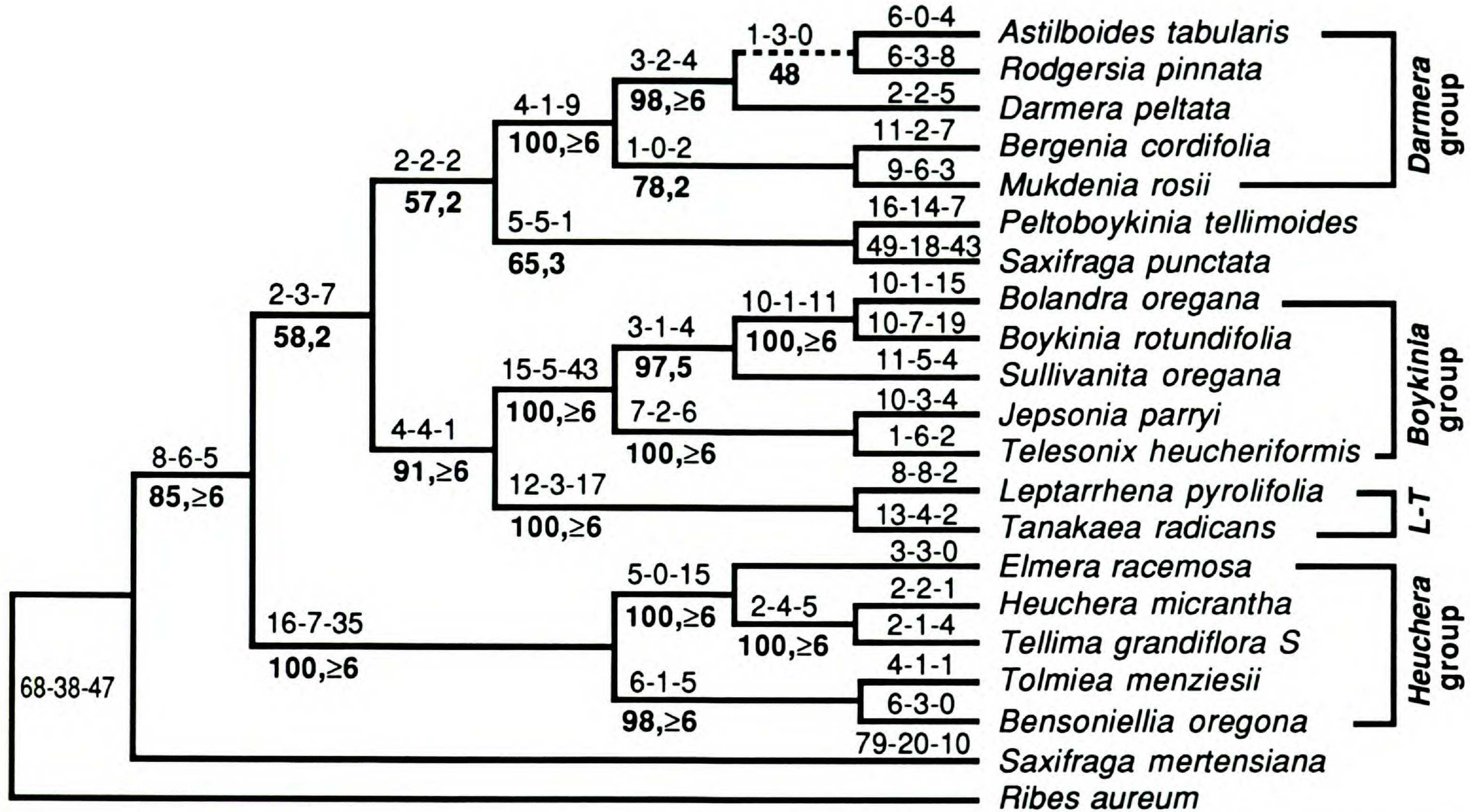


FIGURE 3. One of two most parsimonious trees for Saxifragaceae s. s. resulting from analysis of combined *mat*K and *rbc*L sequences and cpDNA restriction site data for 21 taxa (combined matrix-1). Base substitutions (ACCTRAN) are indicated above branches and are partitioned by data set in the following order: *mat*K-*rbc*L-cpDNA restriction sites. Bootstrap and decay values are indicated below branches, respectively. The dashed line represents the single branch that is not supported by both most parsimonious trees; the decay value is zero and is not indicated following the bootstrap value below the branch.

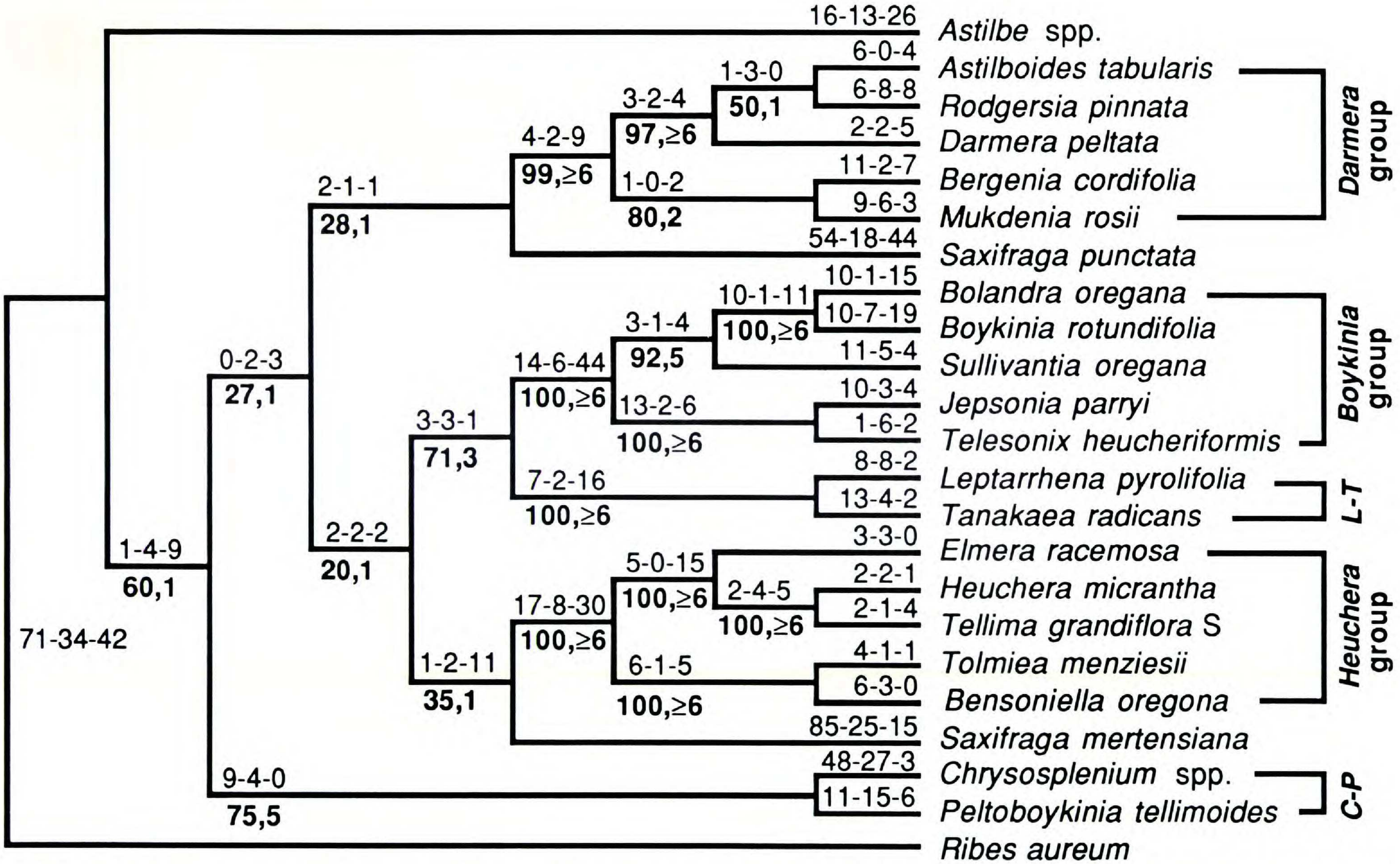


FIGURE 4. Single most parsimonious tree for Saxifragaceae s. s. resulting from analysis of combined *mat*K and *rbc*L sequences and cpDNA restriction site data for 23 taxa (combined matrix-2). Base substitutions (ACCTRAN) are indicated above branches and are partitioned by data set in the following order: *mat*K-*rbc*L-cpDNA restriction sites. Bootstrap and decay values are indicated below branches, respectively.

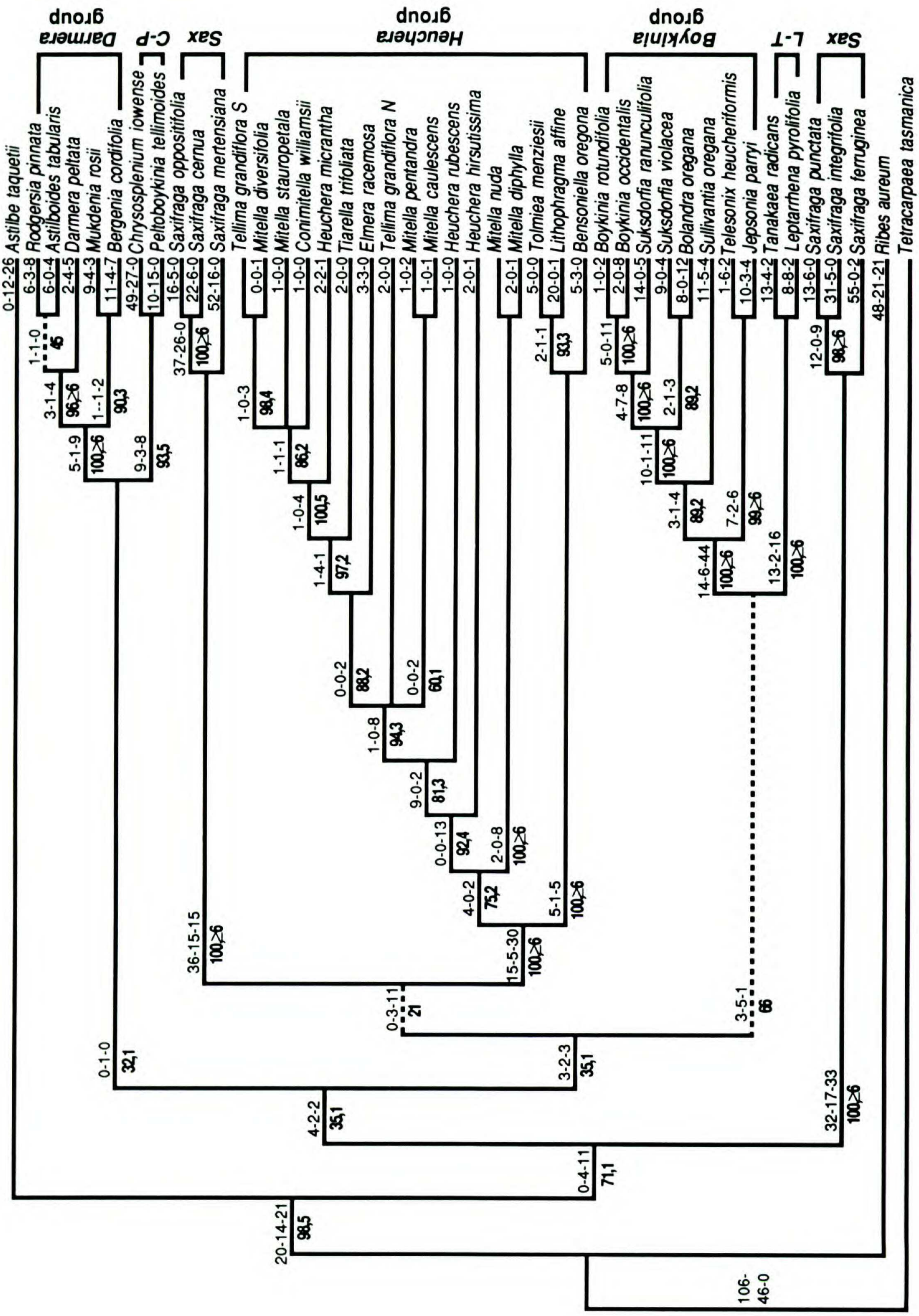

Darmera group
C-P
Sax
Heuchera group
Boykinia group
L-T
Sax
Astilbe taquetii
Rodgersia pinnata
Astilboides tabularis
Darmera peltata
Mukdenia rosii
Bergenia cordifolia
Chrysosplenium iowense
Peltoboykinia tellimoides
Saxifraga oppositifolia
Saxifraga cernua
Saxifraga mertensiana
Tellima grandiflora S
Mitella diversifolia
Mitella stauropetala
Conimitella williamsii
Heuchera micrantha
Tiarella trifoliata
Elmera racemosa
Tellima grandiflora N
Mitella pentandra
Mitella caulescens
Heuchera rubescens
Heuchera hirsutissima
Mitella nuda
Mitella diphylla
Tolmiea menziesii
Lithophragma affine
Bensoniella oregona
Boykinia rotundifolia
Boykinia occidentalis
Suksdorfia ranunculifolia
Suksdorfia violacea
Bolandra oregana
Sullivantia oregana
Telesonix heucheriformis
Jepsonia parryi
Tanakaea radicans
Leptarrhena pyrolifolia
Saxifraga punctata
Saxifraga integrifolia
Saxifraga ferruginea
Ribes aureum
Tetracarpaea tasmanica

Fig. 4). Matrix-3 yielded 24 most parsimonious trees of 1606 steps with a CI of 0.727 (autapomorphies retained; Fig. 5). Our analyses of these three matrices all demonstrated the presence of the same well-supported groups of genera (i.e., the *Boykinia* group, the *Darmera* group, the *Heuchera* group, and *Leptarrhena/Tanakaea*), groups also found in independent analyses of *mat*K sequences (Fig. 2) and *rbc*L sequences and cpDNA restriction sites (Soltis et al., 1993). In contrast, relationships among these well-supported groups of genera are poorly resolved, and there is considerable disagreement in relationships at basal branches among the shortest trees resulting from each of these three combined matrices (Figs. 3–5).

Gilia. Parsimony analysis of *mat*K sequences for *Gilia* yielded a single island of 36 most parsimonious trees each of 418 steps (Fig. 6). The consistency index excluding autapomorphies for these trees is 0.688, which is higher than the expected value for 31 taxa (Sanderson & Donoghue, 1989; Table 2).

VARIABILITY ANALYSES

Comparisons between genes. The average number of nucleotide differences per site in pairwise comparisons of entire (1518–1521 bp) *mat*K sequences for *Bensoniella oregona*, *Saxifraga integrifolia*, and *Sullivantia oregana* is 0.068. This value is 3.2 times greater than the average 0.021 nucleotide differences per site observed in pairwise comparisons of entire *rbc*L sequences for these same species. Comparisons of partial (747–845 bp) *mat*K sequences with *rbc*L (1392 bp) among gymnosperms (*Pinus–Cunninghamia*, *Cunninghamia–Widdringtonia*, *Widdringtonia–Juniperus*, and *Juniperus–Microbiota*) similarly reveals an average 3.4 times greater level of nucleotide differences per site in *mat*K than in *rbc*L (P. Gadek & C. Quinn, pers. comm.).

Extending pairwise comparisons of nucleotide variability to 25 saxifragaceous taxa for which 1078 bp of *mat*K and all of *rbc*L has been sequenced gives an average 3.1-fold greater number of nucleotide differences per site in *mat*K as compared to *rbc*L. The lowest ratio (1.2 : 1) of nucleotide differences per site in this comparison occurred in two pairwise comparisons between members of the *Heuchera* group of genera (*Elmera–Tellima* and *Elmera–Heuchera*) that are only weakly differentiated by sequence data (Figs. 2 and 5). The highest ratio (6.2 : 1) of *mat*K to *rbc*L nucleotide differences per site also occurred within a single group of genera, the *Boykinia* group, between the well-differentiated genera *Bolandra* and *Jepsonia* (Figs. 2 and 5). When the percentage of base positions that are variable is compared across these 25 taxa, *mat*K has 3.2 times as many variable base positions and 2.7 times as many potentially informative characters as does *rbc*L (Table 3). The greater level of variation in *mat*K as compared to *rbc*L also extends to amino acids. Whereas 5% of amino acid positions are variable in *rbc*L, in *mat*K the figure is 59% (Table 3, Fig. 7). Although regions of high variability and regions that are more conserved are apparent in *mat*K sequences, variable sites appear to be fairly uniform in distribution throughout the 5′ portion of *mat*K that we have sequenced (Fig. 7). Character-state reconstructions over a most parsimonious tree for the *mat*K and *rbc*L data sets reveal that the variability in nucleotide substitutions is partitioned much more evenly in *mat*K with regard to transition : transversion and codon position ratios compared to *rbc*L (Table 3).

The average number of nucleotide differences per site in pairwise comparisons of *mat*K sequences (two-thirds complete) from *Coriandrum*, *Griselinia*, and *Hedera* are 1.3-fold greater than the average number observed in comparisons of entire *ndh*F sequences for these same species. The information content per gene is comparable between *ndh*F and *mat*K, however, given that *ndh*F is approximately 1.4 times as long as complete *mat*K sequences. In contrast, pairwise comparisons between *mat*K sequences from *Gilia leptalea*, *G. scopulorum*, and *G. splendens* and sequences from the combined nuclear ITS-1 and ITS-2 regions indicate an average 1.9-fold greater level of nucleotide differences per site in the ITS regions than in *mat*K. The information content of *mat*K is much greater than that of ITS, however, given that *mat*K

←

FIGURE 5. One of 24 most parsimonious trees for Saxifragaceae s. s. resulting from analysis of combined *mat*K and *rbc*L sequences and cpDNA restriction site data for 43 taxa (combined matrix-3). Base substitutions (ACCTRAN) are indicated above branches and are partitioned by data set in the following order: *mat*K-*rbc*L-cpDNA restriction sites. Bootstrap and decay values are indicated below branches, respectively. Dashed lines represent branches that are not supported by all most parsimonious trees; in these instances the decay value is zero and is not indicated following the bootstrap value below the branch.

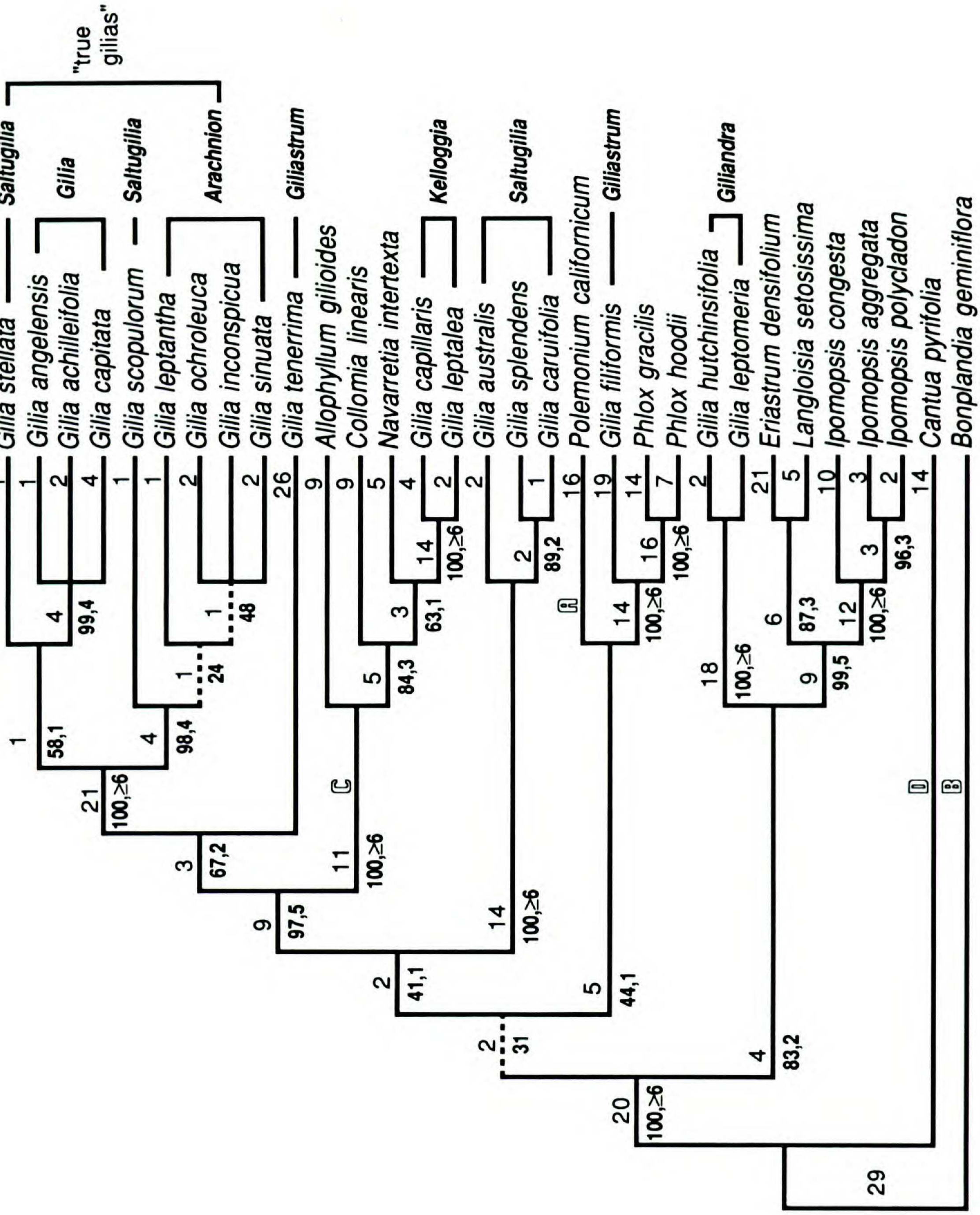

"true gilias"
Gilia stellata
Gilia angelensis
Gilia achilleifolia
Gilia capitata
Gilia scopulorum
Gilia leptantha
Gilia ochroleuca
Gilia inconspicua
Gilia sinuata
Gilia tenerrima
Allophyllum gilioides
Collomia linearis
Navarretia intertexta
Gilia capillaris
Gilia leptalea
Gilia australis
Gilia splendens
Gilia caruifolia
Polemonium californicum
Gilia filiformis
Phlox gracilis
Phlox hoodii
Gilia hutchinsifolia
Gilia leptomeria
Eriastrum densifolium
Langloisia setosissima
Ipomopsis congesta
Ipomopsis aggregata
Ipomopsis polycladon
Cantua pyrifolia
Bonplandia geminiflora
Saltugilia
Gilia
Saltugilia
Arachnion
Giliastrum
Kelloggia
Saltugilia
Giliastrum
Giliandra
A
C
D
B

TABLE 2. Summary of descriptive measures and indices of variation in our *mat*K sequence matrices for Saxifragaceae s.s. (45 taxa) and *Gilia* (31 taxa).

Measure	Saxifragaceae s.s.	*Gilia*
Number of characters	1078	1083
Percent of characters variable	44	26
Percent of characters potentially informative	25	16
Consistency index (CI)	0.735	0.785
CI excluding autapomorphies	0.583	0.688
Expected CI excluding autapomorphies (Sanderson & Donoghue, 1989)	0.341	0.423
Retention index (RI)	0.792	0.871
Homoplasy excess ratio (HER)	0.727	0.827
Transition : transversion ratio	1 : 1.09	1 : 1.05
Codon position ratio	1.17 : 1 : 1.58	1.44 : 1 : 1.87

is approximately 3.1 times longer than the ITS regions combined.

Variation in the noncoding trnK intron. To date, phylogenetic analyses using the *trn*K region have concentrated on *mat*K; however, the *trn*K intron regions flanking *mat*K (Fig. 1) may also have phylogenetic potential. Although we have not used the intron regions between the 5′ *trn*K exon and *mat*K (5′ intron region) and *mat*K and the 3′ *trn*K exon (3′ intron region) for phylogenetic analyses, we have sequenced the major portion of these regions for *Bensoniella oregona*, *Saxifraga integrifolia*, and *Sullivantia sullivantii*. The 5′ intron region is approximately 720 bp with an average 0.059 nucleotide differences per site in pairwise comparisons of these species. The 3′ region is approximately 200 bp with an average 0.116 nucleotide differences per site. For proper alignment of these three sequences, each region required the insertion of five gaps of one to nine bp. The observed variability in these regions flanking *mat*K is slightly less than *mat*K itself for the 5′ flanking region but considerably higher for the 3′ flanking region. Both regions may provide additional, useful phylogenetic information.

Variation in Saxifragaceae s. s. and Gilia matK matrices. As reconstructed over one of their most parsimonious trees (Figs. 2 and 6, respectively), the number of transitions was essentially equivalent to the number of transversions observed in both the *mat*K data set for Saxifragaceae s. s. and *Gilia* (Table 2). As expected, substitutions at the second codon position were less frequent than those at the first or third positions. However, the numbers of substitutions at the first and second codon positions were only slightly lower than the number observed at the third position in both the Saxifragaceae s. s. and the *Gilia mat*K data sets (Tables 2 and 3). In contrast, in *rbc*L the number of third position substitutions is much higher than the numbers of substitutions at the first and second codon positions, as is typical of protein-coding regions with strong functional constraints (Table 3; Donoghue et al., 1992; Smith et al., 1993; see also Steele & Vilgalys, 1994).

The randomization test of Archie (1989a) indicates that variation in the Saxifragaceae s. s. and *Gilia mat*K data sets is substantially non-randomly structured. The most parsimonious tree length (Fig. 2; 842 steps) obtained from analysis of the *mat*K sequences for Saxifragaceae s. s. is far removed from the range (1188 to 1213 steps) and mean (1200 steps) of the most parsimonious tree lengths derived from 100 random permutations of these data. Likewise, the most parsimonious tree length (418 steps) obtained from analysis of the *mat*K sequences for *Gilia* is far removed from the range (838 to 859 steps) and mean (847 steps) of the most parsimonious tree lengths derived from 100 random permutations of the *Gilia* data matrix. Because HER (Archie, 1989b) measures congruence among characters and departure from randomness (i.e., HER equals one when data are com-

←

FIGURE 6. One of 36 most parsimonious trees from analysis of *mat*K sequences for *Gilia* and related genera. Base substitutions (ACCTRAN) are indicated above branches. Bootstrap and decay values are indicated below branches, respectively. Dashed lines represent branches that are not supported by all most parsimonious trees; in these instances the decay value is zero and thus is not indicated following the bootstrap value below the branch. Letters (A–D) denote the distribution of specific indels referenced in Appendix 2.

TABLE 3. Comparison of sequence variation between *mat*K and *rbc*L for the same suite of 25 saxifragaceous taxa. Taxa included in this comparison are indicated in Appendix 1.

Comparison	*mat*K	*rbc*L
Number of characters	1078	1398
Percent of characters variable	38	12
Percent of characters potentially informative	16	7
Percent of amino acid positions variable	59	5
Number of most parsimonious trees	3	24
Consistency index (CI)	0.790	0.712
CI excluding autapomorphies	0.635	0.580
Retention index (RI)	0.736	0.699
Transition : transversion ratio	1 : 1.06	1.41 : 1
Codon position ratio	1.22 : 1 : 1.57	1.26 : 1 : 6.17
RI by codon position	0.76; 0.71; 0.73	0.57; 0.63; 0.75

pletely congruent and approaches zero as data approach randomness in the distribution of character states), we use HER as an indicator of the degree of hierarchical structure present in our data matrices. Values of HER are 0.727 and 0.827 for our Saxifragaceae s. s. and *Gilia mat*K data sets, respectively. These values indicate that homoplasy and departure from hierarchical structure in these *mat*K data are approximately 27.3% (in Saxifragaceae s. s.) and 17.3% (in *Gilia*) of that present in randomly structured data sets possessing the same character-state distributions as our original matrices.

DISCUSSION

PHYLOGENETIC ANALYSES

Saxifragaceae s. s. Several recent studies have presented phylogenetic trees based on *rbc*L sequences and cpDNA restriction sites for members of Saxifragaceae s. s. In addition, the implication of these cpDNA-based phylogenetic trees regarding trends in chemical, morphological, and cytological evolution within the family have been discussed. Herein we focus on comparisons between previous phylogenetic trees based on *rbc*L sequences and cpDNA restriction site variation (Morgan & Soltis, 1993; Soltis et al., 1993) and relationships suggested by *mat*K sequences when analyzed both independently and combined with the other two data sets.

Strong support for the monophyly of Saxifragaceae s. s. has been demonstrated by an extensive analysis of *rbc*L sequences representing a diverse array of dicots (Morgan & Soltis, 1993), as well as with *rpl*2 intron data (Downie et al., 1991). Parsimony analysis of 1078 bp of *mat*K sequence for 45 taxa also reveals a well-supported Saxifragaceae s. s. (Fig. 2). The members of Saxifragaceae s. s. share seven unique base substitutions, are united in all trees found during 100 bootstrap replications, and form a clade in all trees up to at least six steps longer than the most parsimonious trees (Fig. 2), relative to the two outgroup taxa chosen.

Within Saxifragaceae s. s., several lineages and groups of genera (i.e., the *Boykinia*, *Heuchera*, and *Leptarrhena*/*Tanakaea* groups, and two lineages of *Saxifraga*) are well supported by *mat*K sequences, whereas other groups of genera (i.e., the *Darmera* and *Chrysosplenium*/*Peltoboykinia* groups) are less well supported but consistently united in all most parsimonious trees (Fig. 2). Although we have increased both the number of nucleotides and the number of taxa in the *mat*K matrix by almost 50% over an earlier analysis (Johnson & Soltis, 1994), relationships among all of the major lineages noted above remain unresolved. Low support for those branches uniting the major groups of genera was also evident in independent analyses of *rbc*L sequences and cpDNA restriction sites (Soltis et al., 1993) and in the combined analyses of all three cpDNA data sets (Figs. 3–5). Whereas the inability of these three cpDNA data sets to elucidate relationships among the major groups of genera in Saxifragaceae s. s. may seem unsatisfactory from both a cladistic and a taxonomic standpoint, it nonetheless is significant in that it supports the occurrence of a rapid radiation early in the evolutionary history of Saxifragaceae s. s., an event also suggested by other lines of evidence (e.g., host preferences of *Puccinia* rusts; Savile, 1975).

In contrast to the lack of support among the major groups of genera, relationships within some of these groups are well supported by *mat*K sequences. Phylogenetic relationships among members of the *Boykinia* group are, for example, most-

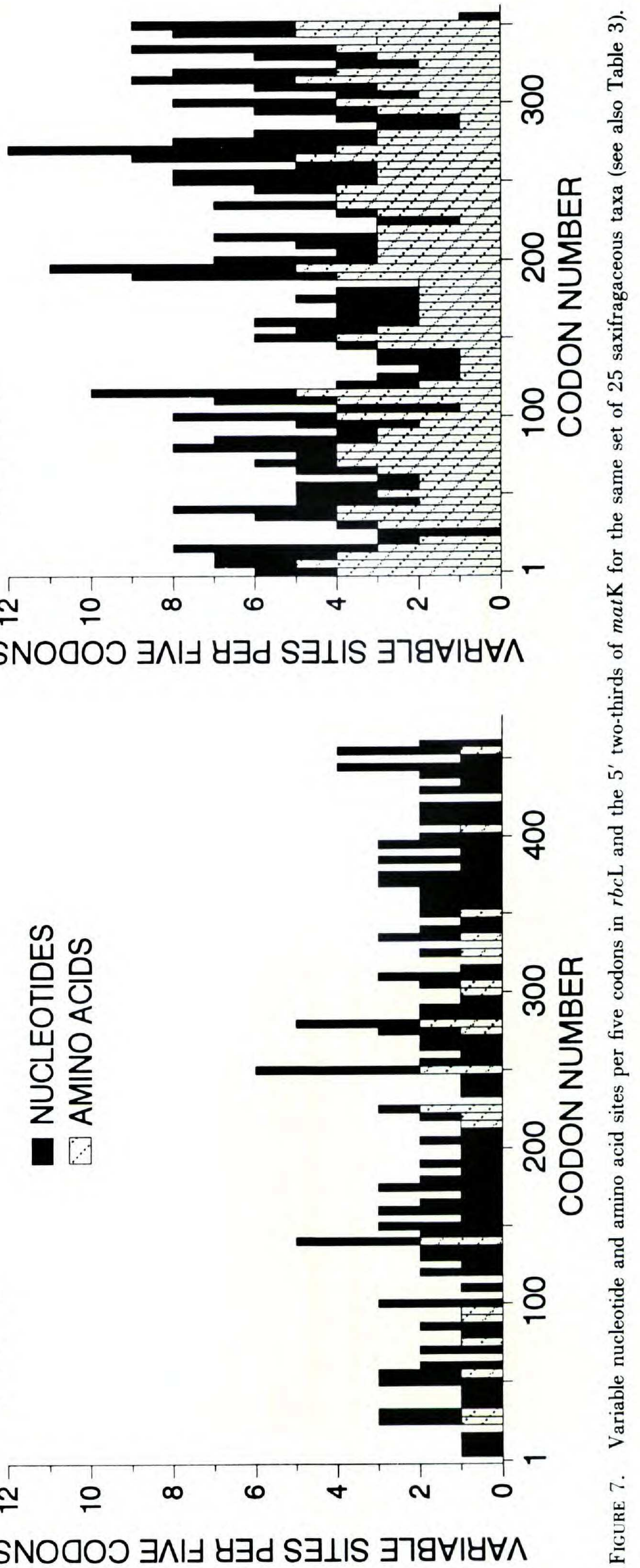


FIGURE 7. Variable nucleotide and amino acid sites per five codons in *rbc*L and the 5′ two-thirds of *mat*K for the same set of 25 saxifragaceous taxa (see also Table 3).

ly well defined and strongly supported by bootstrap values of 100 and decay values of four or higher (Fig. 2). Significantly, the degree of support for generic- and species-level relationships within the *Boykinia* group based on *mat*K sequences is comparable to that obtained via analysis of cpDNA restriction sites (Soltis et al., 1993); these same strongly supported relationships are also seen in the analysis of the three combined cpDNA data sets (Fig. 5). Trees based on *mat*K sequences (Fig. 2) agree with those obtained from analysis of cpDNA restriction sites: in recognizing a strong relationship between *Telesonix* and the enigmatic *Jepsonia*; in suggesting that *Sullivantia* is sister to *Bolandra*, *Boykinia*, and *Suksdorfia*; and in indicating the polyphyly of *Suksdorfia*.

Phylogenetic analysis of *mat*K sequences not only indicates that *Saxifraga* is polyphyletic, comprising two well-supported lineages, but also reveals considerable differentiation among species within each of these two lineages (Fig. 2). *Saxifraga integrifolia*, *S. ferruginea*, *S. mertensiana*, and *S. punctata* (= *S. nelsoniana*) have traditionally been placed in section *Micranthes* (Engler & Irmscher, 1916–1919). Gornall (1987), however, considered *S. mertensiana* so distinctive morphologically that he placed it in its own section, *Heterisia*. Significantly, *S. mertensiana* is well removed from these three species of section *Micranthes* in all most parsimonious trees derived from analysis of *mat*K sequences (Fig. 2) and differs from these species by a minimum of 135 nucleotide differences and 2 indels. The distant relationship of *S. mertensiana* to members of section *Micranthes* (Fig. 2) is also suggested by independent analyses of *rbc*L sequences and cpDNA restriction sites (Soltis et al., 1993), as well as analyses of the combined cpDNA data sets (Figs. 3–5). Additionally, the sister relationship suggested by analysis of *mat*K between *S. mertensiana* and *S. cernua/S. oppositifolia* (of sections *Saxifraga* and *Porphyrion*, respectively) is also suggested by *rbc*L sequences (Soltis et al., 1993). Given the large number (35) of base substitutions uniting *S. mertensiana* with *S. cernua/S. oppositifolia* and an even larger number (52) of autapomorphies in *S. mertensiana*, we cannot discount the possibility that this relationship is an artifact produced by attraction of long branches during parsimony analysis (Felsenstein, 1978). Nonetheless, the level of differentiation in *mat*K observed among the small sample of *Saxifraga* species included herein (Fig. 2; range of 44 to 162 base substitutions in pairwise comparisons) suggests that additional comparative *mat*K sequencing within *Saxifraga* should be extremely useful at defining lineages and relationships among many of the 300 species that are presently recognized in this morphologically diverse genus.

A fourth well-supported, albeit small, lineage in Saxifragaceae s. s. comprises the sister genera *Leptarrhena* and *Tanakaea* (Fig. 2). Previous analyses of *mat*K sequences (Johnson & Soltis, 1994) and *rbc*L sequences (Soltis et al., 1993), as well as the analysis of the three combined cpDNA data sets (Figs. 3–5), not only support this relationship, but also suggest that these two genera are sister to the *Boykinia* group. This relationship to the *Boykinia* group is generally not strongly supported, however, and is not revealed in all most parsimonious trees from analysis of our expanded *mat*K data set (Fig. 2), analysis of combined cpDNA data matrix-3 (Fig. 5), or analysis of cpDNA restriction site data (Soltis et al., 1993). Matrix-1 of the combined cpDNA data sets (Fig. 3) provides the strongest support for a sister relationship between *Leptarrhena/Tanakaea* and the *Boykinia* groups (91% bootstrap and decay value ≥6). However, the small number of taxa in this matrix does not include what may be critical taxa for defining the true affinities of *Leptarrhena/Tanakaea*. For example, matrix-2 of the combined cpDNA data sets includes just two additional genera, *Astilbe* and *Chrysosplenium*, yet support for *Leptarrhena/Tanakaea* as sister to the *Boykinia* group declines to 71% bootstrap and a decay value of 3 in analyses of this data set (Fig. 4).

The *Heuchera* group of genera is strongly supported by *mat*K sequences, and several lineages are differentiated within this group. As was observed in an earlier analysis of cpDNA restriction site data (Soltis et al., 1991), *Heuchera* and *Mitella* appear polyphyletic with species located in several independent lineages. Both cpDNA restriction sites and *mat*K sequences also provide strong support for a clade comprising *Bensoniella*, *Lithophragma*, and *Tolmiea*. *mat*K and cpDNA restriction site data sets do not concur, however, in the affinities of *Bensoniella*, *Lithophragma*, and *Tolmiea* within the *Heuchera* group. Analysis of cpDNA restriction sites suggests that *Bensoniella*, *Lithophragma*, and *Tolmiea* are the sister to all other members of the *Heuchera* group. In contrast, analysis of *mat*K sequences suggests that *Bensoniella*, *Lithophragma*, and *Tolmiea* are the sister to *Conimitella*, *Elmera*, *Tellima*, *Tiarella*, and most species of *Heuchera* and *Mitella*; all of these taxa then appear as the sister to *Heuchera hirsutissima*, *Mitella diphylla*, and *M. nuda* (Fig. 2). The latter three species are well differentiated from other species of *Heuchera* and *Mitella* in the

cpDNA restriction site study (Soltis et al., 1991). The unusual position of these three species in the *Heuchera* group suggested by *mat*K sequences requires further investigation.

Analysis of *mat*K sequences provides support for *Chrysosplenium* and *Peltoboykinia* as sister genera (Fig. 2), a relationship that is also suggested by *rbc*L sequences but not by all shortest trees based on cpDNA restriction sites (Soltis et al., 1993). Not only is the degree of bootstrap support for this relationship high (88%), but these two genera are also united in all trees up to 4 steps longer than the most parsimonious tree and receive similar support in analyses of the combined cpDNA data sets (Figs. 4, 5). The sister relationship between these two genera is significant because both genera are morphologically distinctive and their affinities are problematic. *Chrysosplenium* is particularly well defined morphologically and contains a large number of species. It is thus noteworthy that the branch uniting the two *Chrysosplenium* species is among the longest observed on the *mat*K phylogenetic tree (Fig. 2), with 47 substitutions and two unique indels uniting these species. Comparison of a few partial sequences from other species of *Chrysosplenium* (D. Soltis, unpublished) suggests that *mat*K sequences will be useful for elucidating relationships within this genus.

The *Darmera* group of genera is also recognized based on analysis of *mat*K sequences, although this group is not as strongly supported (76% bootstrap, decay value of 2; Fig. 2) as are some of the other major groups of genera in Saxifragaceae s. s. The *Darmera* group is, however, well supported by cpDNA restriction sites (99% bootstrap, ≥6 steps of decay; Soltis et al., 1993) and also appears in our analyses of the three combined cpDNA data sets (Figs. 3–5). In contrast, *rbc*L sequences fail to recognize the *Darmera* group, although this appears to be the result of the small number of base substitutions in *rbc*L at this level of analysis rather than indicative of a strongly supported opposing view of relationships (Soltis et al., 1993). As within other major groups of genera in Saxifragaceae, *mat*K and cpDNA restriction sites again provide comparable pictures of relationships within the *Darmera* group. Both cpDNA restriction sites and *mat*K sequences indicate that *Bergenia* and *Mukdenia* are sister taxa and that *Astilboides*, *Darmera*, and *Rodgersia* are closest relatives.

Gilia. In a recent analysis of 661 bp of *mat*K sequence for 20 polemoniaceous taxa, *Gilia* appeared monophyletic (Steele & Vilgalys, 1994); however, Steele and Vilgalys included only four species of *Gilia* from two sections, *Arachnion* and *Gilia*. Our broader analysis of *mat*K sequences from species representing all six sections of *Gilia* reveals that species of *Gilia* are scattered throughout the temperate radiation of Polemoniaceae (Fig. 6). Species of *Gilia* appear in several well-separated lineages that also include members of virtually all of the other temperate genera included in this analysis (i.e., *Allophyllum*, *Collomia*, *Eriastrum*, *Ipomopsis*, *Langloisia*, *Navarretia*, *Phlox*, and *Polemonium*). Thus, our analysis clearly indicates that *Gilia* is polyphyletic. Although few of the species of *Gilia* included in the *mat*K analysis were identical to those sampled in a recent ITS sequence analysis of Polemoniaceae (Fig. 8; Porter, 1993), the sectional coverage of *Gilia* in the two studies is comparable. Both *mat*K and ITS phylogenetic trees agree in suggesting similar relationships among sections of *Gilia* and allied genera (compare Figs. 6 and 8). Furthermore, relationships that are only weakly supported by *mat*K sequence data, such as the branching patterns among basal nodes (Fig. 6), are also only weakly supported by ITS sequences (Fig. 8).

Analyses of *mat*K sequences strongly support a clade of "true gilias" comprising sections *Gilia*, *Arachnion*, and the species *G. scopulorum* and *G. stellata* of section *Saltugilia* (Fig. 6). A close relationship among sections *Arachnion*, *Gilia*, and *Saltugilia* has been previously suggested (Grant, 1954, 1959; Grant & Grant, 1956a, b). Furthermore, although both *G. scopulorum* and *G. stellata* have been allied with the *G. splendens* group of section *Saltugilia* (Grant & Grant, 1954, 1956b; Grant, 1959; Day, 1993b), the distinctness of the former species with regard to the *G. splendens* group has also been recognized (Grant & Grant, 1954, 1956b). ITS sequences also define a clade of true gilias similar in composition to that defined by *mat*K sequences. However, in the ITS analysis *G. stellata* and *G. scopulorum* are united in a single clade as sister to both sections *Gilia* and *Arachnion* (Fig. 8), rather than placed apart as suggested by *mat*K sequences (Fig. 6).

Gilia tenerrima (section *Giliastrum*) appears as a weakly supported sister to the true gilias in analyses of both *mat*K (Fig. 6) and ITS (Fig. 8) sequences. The morphology of *G. tenerrima* is unique among gilias, and its affinities have been sought among species currently recognized in section *Kelloggia* (e.g., Mason & Grant, 1948; Grant, 1959) or section *Giliastrum* (e.g., Grant & Grant, 1954; Day, 1993a). The polyphyly of section *Giliastrum* is indicated by *mat*K sequences based on the inclusion of only one other species of section *Gi-*

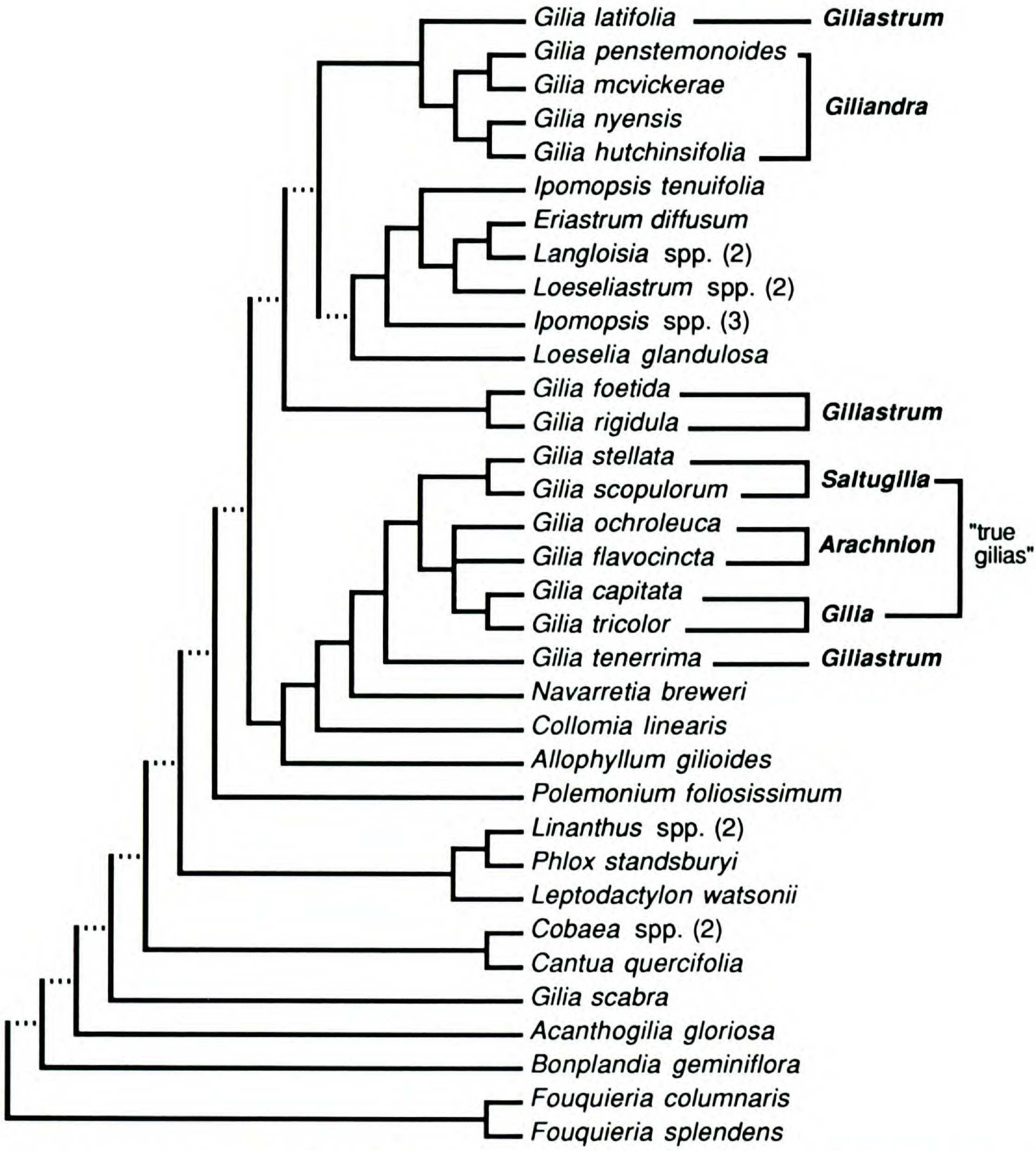


FIGURE 8. Strict consensus of two most parsimonious trees for Polemoniaceae resulting from parsimony analysis of nuclear rDNA ITS sequences (length = 1074; CI = 0.47). Poorly supported branches, as indicated by bootstrap, jackknife, decay, and skewness (Hillis & Huelsenbeck, 1992) analyses, are represented by broken lines. A few terminal branches have been collapsed to emphasize relationships discussed herein; in these cases, the value in parentheses to the right of the generic name represents the number of species that appear along the branch. Data analysis and tree courtesy of Porter (1993).

liastrum, *G. filiformis*, in this analysis. *Gilia filiformis* is strongly allied with a lineage including *Phlox* in the *mat*K analysis (Fig. 6) and is well separated from *G. tenerrima* by 65 bases in a pairwise comparison of nucleotide differences. The polyphyly of section *Giliastrum* is also strongly suggested by ITS sequences (Fig. 8).

A well-supported sister lineage to *G. tenerrima* and the clade of true gilias is a strongly supported group comprising *Allophyllum*, *Collomia*, *Navarretia*, and species of *Gilia* section *Kelloggia*. These taxa share a 6-bp deletion and 10 unambiguous base substitutions on the *mat*K phylogenetic tree (Fig. 6). *Gilia capillaris* and *G. leptalea* (section

Kelloggia; Mason & Grant, 1948; Day, 1993a) form a well-defined lineage within this group and are differentiated from their sister genera by eight unambiguous base substitutions. These two species have, until recently (Day, 1993a), been recognized as a distinct species group in section *Saltugilia* (Grant & Grant, 1954; Grant, 1959). Analyses of *mat*K sequences thus support the recognition of section *Kelloggia*, although the close relationship of this section to *Allophyllum*, *Collomia*, and *Navarretia* has not been considered in recent classifications. It is noteworthy, however, that Gray (1873) and Brand (1907) placed *G. leptalea* and *G. sinistra*, respectively, in *Collomia* (*G. sinistra* is also in section *Kelloggia* but was not included in this analysis of *mat*K sequences). Species from section *Kelloggia* were not included in Porter's (1993) ITS analysis, and ITS sequences place *Navarretia*, *Collomia*, and *Allophyllum* as progressively basal sisters to *G. tenerrima* and the true gilias rather than united as a single lineage as suggested by *mat*K sequences (Fig. 6). This apparent discrepancy between the ITS and *mat*K trees may simply be a result of inappropriate rooting of the ITS sequence clade that includes *Allophyllum* and the true gilias due to the inclusion of very divergent sequences in the ITS matrix (M. Porter, pers. comm.). For example, the *Allophyllum*-true gilia clade in the ITS tree (Fig. 8) could be rerooted at the midpoint of the branch between *G. tenerrima* and *Navarretia* to provide essentially the same view of relationships depicted by the *mat*K sequence tree (Fig. 6).

The core members of section *Saltugilia*, *G. splendens*, *G. australis*, and *G. caruifolia*, were also not included in Porter's (1993) ITS sequence analysis but are strongly supported by *mat*K data as a unique lineage well separated from all other lineages of *Gilia* sequenced to date. A close relationship among these three species was first formally recognized by Grant & Grant (1954). However, the great divergence between this lineage and the true gilia clade has not been previously considered. We are currently further investigating the affinities of the *G. splendens* group because its placement as sister to the *Allophyllum* and true gilia clades in the *mat*K analysis is only weakly supported (Fig. 6).

Finally, this analysis of *mat*K sequences recognizes *G. hutchinsifolia* and *G. leptomeria* (both of section *Giliandra*) as a well-supported clade that is also well separated from other gilias on the most parsimonious tree (Fig. 6). A clade including *G. hutchinsifolia* is similarly well removed from other gilias in analyses of ITS sequences (Fig. 8). Furthermore, both *mat*K and ITS sequences concur in placing *G. hutchinsifolia* and related species as sister to *Eriastrum*, *Langloisia*, and *Ipomopsis* (Fig. 6), although this relationship is not strongly supported in either analysis.

Polyploid origin of Saxifraga osloensis. The ability of *mat*K sequences to reveal even fine-scale relationships in taxonomic groups is illustrated by an example from the genus *Saxifraga*. *Saxifraga osloensis* is a tetraploid of evolutionary interest because it has been proposed by some as the only saxifrage that can confidently be considered an alloploid, although its parentage is controversial (reviewed in Webb & Gornall, 1989). Furthermore, it is also considered to be a classic example of a species of postglacial allopolyploid origin. Knaben (1954) hypothesized that *S. osloensis* is derived from two closely related diploid species, *S. tridactylites* and *S. adscendens*. C. Brochmann (pers. comm.), in contrast, has suggested that *S. osloensis* may be an autopolyploid derived from only *S. tridactylites*. Webb & Gornall (1989) note, however, that morphologically, *S. osloensis* more closely resembles *S. adscendens*. A comparison of only 750 bp of *mat*K sequence shows that *S. osloensis* is identical in sequence to *S. adscendens*, and the two species differ from *S. tridactylites* by 11 nucleotide differences. Thus, either *S. osloensis* is an autoploid derived from *S. adscendens*, or if an alloploid, *S. adscendens* was the maternal parent.

VARIABILITY

Indels. The Saxifragaceae s. s. and *Gilia* data sets suggest that a *mat*K sequence matrix with enough taxonomic breadth to show informative base substitutions is also likely to contain indels. For example, the 45-taxon Saxifragaceae s. s. *mat*K sequence matrix contains 10 indels (Fig. 2; Appendix 2), the 31-taxon *Gilia mat*K sequence matrix contains four indels (Fig. 6, Appendix 2), a 76-taxon Apiaceae–Pittosporaceae *mat*K sequence matrix contains 12 indels (Plunkett, 1994), and a seven-taxon Cupressaceae *mat*K sequence matrix contains two indels (P. Gadek and C. Quinn, pers. comm.). All of these indels are small, ranging in size from 3 to 9 bp. Because apparently identical indels may have multiple origins in unrelated taxa (Golenberg et al., 1993), we have not appended these indels as additional characters in the data matrices. However, only one of the 16 indels observed in the above matrices (indel E in the Saxifragaceae; Fig. 2, Appendix 2) appears homoplasious when mapped on trees derived from analyses

of base substitutions alone. Furthermore, given the poor resolution in the placement of *Astilbe* and the possibility of long-branch attraction in the placement of *Saxifraga mertensiana* on the most parsimonious trees for Saxifragaceae s. s. (Fig. 2), even this indel may have had a single origin. Although it has been our experience with *mat*K that indels can usually be aligned with a high degree of confidence if the reading frame is taken into consideration, we recognize that uncertainties in the alignment of indels is a potential source of error in analyses (Ritland & Clegg, 1987). The amount of error is likely to be minimal, however, given that few bases are involved in indels relative to the number of potentially informative characters that are unambiguously aligned.

Base substitutions. Most base substitutions in the Saxifragaceae s. s. and *Gilia mat*K data sets are hierarchically structured as evidenced by the randomization test of Archie (1989a). Furthermore, the homoplasy observed in these large *mat*K data matrices is not only modest (Table 2), but is also evenly distributed on the most parsimonious trees (data not shown). Together with the strong bootstrap support and high decay values for many of the clades present in the phylogenetic trees for Saxifragaceae s. s. and *Gilia* (Figs. 2 and 6, respectively), these observations suggest that both matrices have been constructed within an appropriate range of taxonomic hierarchy. Thus, although a minimum of 85% (in Saxifragaceae s. s.) and 73% (in *Gilia*) of the potentially informative characters in these *mat*K sequence matrices have experienced multiple hits (based on the number of potentially informative characters with 3 or 4 character states), most of these multiple hits convey phylogenetic information. Whereas multiple substitutions per site may be problematic in *mat*K analyses at broad taxonomic levels, they do not appear to be unduly problematic in analyses of intergeneric and interspecific relationships.

PCR AND SEQUENCING PRIMERS

To stimulate the use of *mat*K sequence data in other groups, we have tried to assess the broad applicability of the PCR and sequencing primers described herein. The *trn*K coding regions are very conserved among angiosperms, and PCR primers *trn*K-2R and *trn*K-3914F (Table 1) have been used by ourselves and others to amplify *mat*K from dicots such as Annonaceae, Apiaceae, Asteraceae, Balsaminaceae, Brassicaceae, Cornaceae, Ericaceae, Grossulariaceae, Lauraceae, Magnoliaceae, Malpighiaceae, Malvaceae, Polemoniaceae, Rhamnaceae, Rosaceae, Sarraceniaceae and Saxifragaceae. These primers have also been used to amplify *mat*K from monocots such as Iridaceae, Juncaceae, and Orchidaceae, although the latter family requires twice the standardly employed amount of $MgCl_2$ to achieve successful amplification via PCR (M. Chase, pers. comm.). For routine sequencing of monocots, however, it may be desirable to use Learn's original *trn*K-3914F primer (Table 1) that provides even greater homology to the 5′ *trn*K exon in these plants. Outside of angiosperms, primers *trn*K-3914F and *trn*K-2R have been used successfully to amplify *mat*K from Cupressaceae (P. Gadek & C. Quinn, pers. comm.). The similarity of these amplification primers to published *trn*K sequences of *Pinus* (Lidholm & Gustafsson, 1991) is high for *trn*K-3914F, but somewhat lower for *trn*K-2R. It may therefore be desirable to modify *trn*K-2R for routine use in conifers. Both *trn*K-3914F and *trn*K-2R also exhibit high similarity to a published sequence for *Marchantia* (Umesono et al., 1988) and may be useful for amplifying *mat*K from bryophytes as well. The PCR amplification primers *rps*16-4547F and *psb*A-R are useful as alternatives to primers *trn*K-3914F and *trn*K-2R, respectively, in some circumstances, but these primers may not prove as widely applicable (Fig. 1; Table 1). For example, *rps*16 (Fig. 1) is not present in *Marchantia* (Umesono et al., 1988), and *psb*A (Fig. 1) is duplicated between *rps*16 and the 5′ *trn*K exon in at least some species of *Pinus* (Lidholm & Gustafsson, 1991). More recently, primers *trn*K-253F, *trn*K-710F, and *trn*K-2000R have been used in various combinations with the standard PCR primers described above to obtain shorter fragments that contain *mat*K from taxa in Apicaceae (Plunkett, 1994) and *Saxifraga* that otherwise have yielded poor *mat*K PCR products. For sequencing, primers *mat*K-1235R, *mat*K-1470R, and *mat*K-1412F (Table 1) appear fairly conserved, at least among dicots, and enable the generation of sequences of approximately 1100 contiguous bp of *mat*K beginning at the 5′ end. We have used these sequencing primers successfully in Apiaceae, Brassicaceae, Cornaceae, Ericaceae, Grossulariaceae, Polemoniaceae, and Saxifragaceae. Other sequencing primers are less conserved, have not been broadly tested, or, in the case of primer *mat*K-1168R, the 3′ nine nucleotides are duplicated within the annealing site of *mat*K-1470R, and have annealed at both sites in some Saxifragaceae and Polemoniaceae, but not Apiaceae (G. Plunkett, pers. comm.). Of the sequencing primers we have used in angiosperms, *mat*K-1470R has also worked in the Cupressaceae

(P. Gadek & C. Quinn, pers. comm.) and thus appears to be a good choice for initial sequencing of *mat*K in diverse plant groups. Given the rapid rate of *mat*K sequence evolution, we have found that at least some primers will need to be designed specifically for a given group in order to sequence all of *mat*K. For example, whereas *mat*K-1168R, *mat*K-1235R, and *mat*K-1470R worked well in nearly all members of Saxifragaceae s. s. sequenced for this study, comparative sequencing within just the large genus *Saxifraga* has required the synthesis of three additional sequencing primers (*trn*K 710F, *mat*K-1176F, and *mat*K-1412R; D. Soltis, unpublished). The need to design group specific sequencing primers is likely to be true of any gene as large as *mat*K, however, that also evolves at a rate useful for intrafamilial and intrageneric phylogenetic reconstruction.

Literature Cited

Archie, J. W. 1989a. A randomization test for phylogenetic information in systematic data. Syst. Zool. 38: 239–252.

———. 1989b. Homoplasy excess ratios: New indices for measuring levels of homoplasy in phylogenetic systematics and a critique of the consistency index. Syst. Zool. 38: 253–269.

———. 1990. Homoplasy excess statistics and retention indices: A reply to Farris. Syst. Zool. 39: 169–174.

Brand, A. 1907. Polemoniaceae. *In:* A. Engler (editor), Das Pflanzenreich IV. 250: 1–203.

Bremer, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution 42: 795–803.

Brunsfeld, S. J., P. S. Soltis, D. E. Soltis, P. A. Gadek, C. J. Quinn, D. D. Strenge & T. A. Ranker. 1994. Phylogenetic relationships among the genera of Taxodiaceae and Cupressaceae: Evidence from *rbc*L sequences. Syst. Bot. 19: 253–262.

Chase, M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duvall, R. A. Price, H. G. Hills, Y.-L. Qiu, K. A. Kron, J. H. Rettig, E. Conti, J. D. Palmer, J. R. Manhart, K. J. Sytsma, H. J. Michaels, W. J. Kress, K. G. Karol, W. D. Clark, M. Hedrén, B. S. Gaut, R. K. Jansen, K.-J. Kim, C. F. Wimpee, J. F. Smith, G. R. Furnier, S. H. Strauss, Q.-Y. Xiang, G. M. Plunkett, P. S. Soltis, S. M. Swensen, S. E. Williams, P. A. Gadek, C. J. Quinn, L. E. Eguiarte, E. Golenberg, G. H. Learn, Jr., S. W. Graham, S. C. H. Barrett, S. Dayanandan & V. A. Albert. 1993. Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbc*L. Ann. Missouri Bot. Gard. 80: 528–580.

Conti, E., A. Fischbach & K. J. Sytsma. 1993. Tribal relationships in Onagraceae: Implications from *rbc*L sequence data. Ann. Missouri Bot. Gard. 80: 672–685.

Day, A. G. 1993a. New taxa and nomenclatural changes in *Allophyllum*, *Gilia*, and *Navarretia* (Polemoniaceae). Novon 3: 331–340.

———. 1993b. *Gilia*. Pp. 828–836 *in* J. C. Hickman (editor), The Jepson Manual: Higher Plants of California. Univ. California Press, Berkeley.

Donoghue, M. J., R. G. Olmstead, J. F. Smith & J. D. Palmer. 1992. Phylogenetic relationships of Dipsacales based on *rbc*L sequences. Ann. Missouri Bot. Gard. 79: 333–345.

Downie, S. R., R. G. Olmstead, G. Zurawski, D. E. Soltis, P. S. Soltis, J. C. Watson & J. D. Palmer. 1991. Loss of the chloroplast DNA *rpl*2 intron demarcates six lineages of dicotyledons: Molecular and phylogenetic implications. Evolution 45: 1245–1259.

Doyle, J. J. & J. L. Doyle. 1987. A rapid isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19: 11–15.

Engler, A. & E. Irmscher. 1916–1919. *Saxifraga*. *In:* A. Engler (editor), Das Pflanzenreich IV. 117: 1–709. Engelmann, Leipzig.

Farris, J. S. 1989a. The retention index and the rescaled consistency index. Cladistics 5: 417–419.

———. 1989b. The retention index and homoplasy excess. Syst. Zool. 38: 406–407.

Felsenstein, J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. Syst. Zool. 27: 401–410.

———. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39: 783–791.

——— & H. Kishino. 1993. Is there something wrong with the bootstrap on phylogenies? A reply to Hillis and Bull. Syst. Biol. 42: 193–199.

Gadek, P. & C. J. Quinn. 1993. An analysis of relationships within the Cupressaceae sensu stricto based on *rbc*L sequences. Ann. Missouri Bot. Gard. 80: 581–586.

Giannasi, D. E., G. Zurawski, G. Learn & M. T. Clegg. 1992. Evolutionary relationships of the Caryophyllidae based on comparative *rbc*L sequences. Syst. Bot. 17: 1–15.

Gielly, L. & P. Taberlet. 1994. The use of chloroplast DNA to resolve plant phylogenies: Noncoding versus *rbc*L sequences. Molec. Biol. Evol. 11: 769–777.

Golenberg, E. M., M. T. Clegg, M. L. Durbin, J. Doebley & D. P. Ma. 1993. Evolution of a non-coding region of the chloroplast genome. Molec. Phylogenetics Evol. 2: 52–64.

Gornall, R. J. 1987. An outline of a revised classification of *Saxifraga* L. J. Linn. Soc., Bot. 95: 273–292.

Grant, V. 1954. Genetic and taxonomic studies in *Gilia*. VI. Interspecific relationships in the leafy-stemmed gilias. Aliso 3: 35–49.

———. 1959. Natural History of the Phlox Family. Vol. 1. Systematic Botany. Martinus Nijhoff, The Hague, Netherlands.

——— & A. Grant. 1954. Genetic and taxonomic studies in *Gilia*. VII. The woodland gilias. Aliso 3: 59–91.

——— & ———. 1956a. Genetic and taxonomic studies in *Gilia*. VIII. The cobwebby gilias. Aliso 3: 203–287.

——— & ———. 1956b. Genetic and taxonomic studies in *Gilia*. X. Conspectus of the subgenus *Gilia*. Aliso 3: 297–300.

Gray, A. 1873. Revision of the North American Polemoniaceae. Proc. Amer. Acad. Arts 8: 247–282.

HILLIS, D. M. & J. J. BULL. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42: 182–192.

——— & J. P. HUELSENBECK. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. J. Heredity 83: 189–195.

HOOT, S. B., A. CULHAM & P. R. CRANE. 1995. The utility of *atp*B gene sequences in resolving phylogenetic relationships: Comparison with *rbc*L and 18S ribosomal DNA sequences in the Lardizabalaceae. Ann. Missouri Bot. Gard. 82: 194–208.

JOHNSON, L. A. & D. E. SOLTIS. 1994. *mat*K DNA sequences and phylogenetic reconstruction in Saxifragaceae s. s. Syst. Bot. 19: 143–156.

KÄLLERSJÖ, M., J. S. FARRIS, A. G. KLUGE & C. BULT. 1992. Skewness and permutation. Cladistics 8: 275–287.

KIM, K.-J., R. K. JANSEN, R. S. WALLACE, H. J. MICHAELS & J. D. PALMER. 1992. Phylogenetic implications of *rbc*L sequence variation in the Asteraceae. Ann. Missouri Bot. Gard. 79: 428–445.

KLUGE, A. G. & J. S. FARRIS. 1969. Quantitative phyletics and the evolution of anurans. Syst. Zool. 18: 1–32.

KNABEN, G. 1954. *Saxifraga osloensis* n. sp., a tetraploid species of the *Tridactylites* section. Nytt Magasin for Botanikk 3: 117–138.

KRON, K. A. & M. W. CHASE. 1993. Systematics of the Ericaceae, Empetraceae, Epacridaceae, and related taxa based upon *rbc*L sequence data. Ann. Missouri Bot. Gard. 80: 735–741.

KUMAR, S., K. TAMURA & M. NEI. 1993. MEGA: Molecular Evolutionary Genetics Analysis, version 1.0. The Pennsylvania State University, University Park, Pennsylvania.

LIDHOLM, J. & P. GUSTAFSSON. 1991. A three-step model for the rearrangement of the chloroplast *trn*K-*psb*A region of the gymnosperm *Pinus contorta*. Nucl. Acids Research 19: 2881–2887.

MADDISON, D. R. 1991. The discovery and importance of multiple islands of most-parsimonious trees. Syst. Zool. 40: 315–328.

MADDISON, W. P. & D. R. MADDISON. 1992. MacClade, version 3.01. Sinauer Associates, Sunderland, Massachusetts.

MASON, H. L. & A. D. GRANT. 1948. Some problems in the genus *Gilia*. Madroño 9: 201–220.

MORGAN, D. R. & D. E. SOLTIS. 1993. Phylogenetic relationships among members of Saxifragaceae sensu lato based on *rbc*L sequence data. Ann. Missouri Bot. Gard. 80: 631–660.

———, ——— & K. R. ROBERTSON. 1994. Systematic and evolutionary implications of *rbc*L sequence variation in Rosaceae. Amer. J. Bot. 81: 890–903.

NEUHAUS, H. & G. LINK. 1987. The chloroplast $tRNA^{Lys}$ (UUU) gene from mustard (*Sinapsis alba*) contains a class II intron potentially coding for a maturase-related polypeptide. Current Genetics 11: 251–257.

OLMSTEAD, R. G. & P. A. REEVES. 1995. Evidence for the polyphyly of the Scrophulariaceae based on chloroplast *rbc*L and *ndh*F sequences. Ann. Missouri Bot. Gard. 82: 176–193.

——— & J. A. SWEERE. 1994. Combining data in phylogenetic systematics: An empirical approach using three molecular data sets in the Solanaceae. Syst. Biol. 43: 467–481.

———, B. BREMER, K. M. SCOTT & J. D. PALMER. 1993. A parsimony analysis of the Asteridae sensu lato based on *rbc*L sequences. Ann. Missouri Bot. Gard. 80: 700–722.

PLUNKETT, G. M. 1994. A molecular-phylogenetic approach to the "family-pair dilemma" in Apiales and Cyperales. Ph.D. Dissertation, Washington State University, Pullman.

PORTER, J. M. 1993. Phylogenetic systematics of *Gilia* section *Giliandra* (Polemoniaceae). Pp. 15–97, Chapter 1: Phylogeny of Polemoniaceae based on nuclear ribosomal internal transcribed spacer sequences. Ph.D. Dissertation, University of Arizona, Tucson.

PRICE, R. A. & J. D. PALMER. 1993. Phylogenetic relationships of the Geraniaceae and Geraniales from *rbc*L sequence comparisons. Ann. Missouri Bot. Gard. 80: 661–671.

QIU, Y.-L., M. W. CHASE, D. H. LES & C. R. PARKS. 1993. Molecular phylogenetics of the Magnoliidae: Cladistic analyses of nucleotide sequences of the plastid gene *rbc*L. Ann. Missouri Bot. Gard. 80: 587–606.

RITLAND, K. & M. T. CLEGG. 1987. Evolutionary analysis of plant DNA sequences. Amer. Naturalist 130: S74–S100.

RODMAN, J., R. A. PRICE, K. KAROL, E. CONTI, K. J. SYTSMA & J. D. PALMER. 1993. Nucleotide sequences of the *rbc*L gene indicate monophyly of mustard oil plants. Ann. Missouri Bot. Gard. 80: 686–699.

SANDERSON, M. J. & M. J. DONOGHUE. 1989. Patterns of variation in levels of homoplasy. Evolution 43: 1781–1795.

SAVILE, D. B. O. 1975. Evolution and biogeography of Saxifragaceae with guidance from their rust parasites. Ann. Missouri Bot. Gard. 62: 354–361.

SMITH, J. F., W. J. KRESS & E. A. ZIMMER. 1993. Phylogenetic analysis of the Zingiberales based on *rbc*L sequences. Ann. Missouri Bot. Gard. 80: 620–630.

SOLTIS, D. E. & R. K. KUZOFF. Discordance between molecular and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). Evolution: in press.

———, P. S. SOLTIS, M. T. CLEGG & M. DURBIN. 1990. *rbc*L sequence divergence and phylogenetic relationships in Saxifragaceae sensu lato. Proc. Natl. Acad. U.S.A. 87: 4640–4644.

———, ———, T. G. COLLIER & M. L. EDGERTON. 1991. Chloroplast DNA variation within and among genera of the *Heuchera* group (Saxifragaceae): Evidence for chloroplast transfer and paraphyly. Amer. J. Bot. 78: 1091–1112.

———, D. R. MORGAN, A. GRABLE, P. S. SOLTIS & R. KUZOFF. 1993. Molecular systematics of Saxifragaceae sensu stricto. Amer. J. Bot. 80: 1056–1081.

STEELE, K. P. & R. VILGALYS. 1994. Phylogenetic analyses of Polemoniaceae using nucleotide sequences of the plastid gene *mat*K. Syst. Bot. 19: 126–142.

SWOFFORD, D. L. 1991. PAUP: Phylogenetic Analysis Using Parsimony, version 3.1.1. Illinois Natural History Survey, Champaign.

UMESONO, K., H. INOKUCHI, Y. SHIKI, M. TAKEUCHI, Z. CHANG, H. FUKUZAWA, T. KOHCHI, H. SHIRAI, K. OHYAMA & H. OZEKI. 1988. Structure and organization of *Marchantia polymorpha* chloroplast genome. II. Gene organization of the large single

copy region from *rps*12 to *atpβ*. J. Molec. Biol. 203: 299–331.

WEBB, D. A. & R. J. GORNALL. 1989. A Manual of Saxifrages and their Cultivation. Timber Press, Portland.

WHERRY, E. T. 1940. A provisional key to the Polemoniaceae. Bartonia 20: 14–17.

WOLFE, K. H. 1991. Protein-coding genes in chloroplast DNA: Compilation of nucleotide sequences, data base entries and rates of molecular evolution. Pp. 467–482 *in* I. K. Vasil (editor), Cell Culture and Somatic Cell Genetics of Plants, Vol. 7B. Academic Press, San Diego.

———, C. W. MORDEN & J. D. PALMER. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. Proc. Natl. Acad. U.S.A. 89: 10648–10652.

APPENDIX 1. Species included in the *Gilia matK* sequence matrix (referenced to Fig. 6), the Saxifragaceae s.s. *matK* sequence matrix (referenced to Fig. 2), the three Saxifragaceae s.s. matrices composed of combined cpDNA restriction sites (r. s.) and *matK* and *rbcL* sequences (matrix-1 is referenced to Fig. 3; matrix-2 is referenced to Fig. 4; and matrix-3 is referenced to Fig. 5), and the comparison between *rbcL* and *matK* sequences for 25 species (referenced to Table 3). GenBank accession numbers are reported for all *matK* and *rbcL* sequences, whereas cpDNA restriction site data are located in Soltis et al. (1991) and Soltis et al. (1993).

Family Species	Voucher/citation	Data type	GenBank	Figure/Table
GROSSULARIACEAE				
Ribes aureum Pursh	*Soltis & Soltis 2220*, WS	*matK*	L34153	2, 3, 4, 5/3
	Morgan & Soltis, 1993	*rbcL*	L11204	3, 4, 5/3
	Soltis et al., 1993	r. s.	—	3, 4, 5
Tetracarpaea tasmanica Hook. f.	*Jordon s.n.*, HO	*matK*	L34154	2, 5/3
	Morgan & Soltis, 1993	*rbcL*	L11207	5/3
POLEMONIACEAE				
Allophyllum gilioides (Benth.) Grant & Grant	*Johnson 92012*, WS	*matK*	L34176	6
Bonplandia geminiflora Cav.	*Patterson s.n.*, WS	*matK*	L34179	6
Cantua pyrifolia Juss.	*Patterson s.n.*, WS	*matK*	L34180	6
Collomia linearis Nutt.	*Johnson 92045*, WS	*matK*	L34188	6
Eriastrum densifolium (Benth.) Mason	*Johnson 92090*, WS	*matK*	L34184	6
Gilia achilleifolia Benth.	*Schultz s.n.*, WS	*matK*	L34175	6
Gilia angelensis Grant	*Johnson 92013*, WS	*matK*	L34177	6
Gilia australis (Mason & Grant) Grant & Grant	*Johnson 92021*, WS	*matK*	L34178	6
Gilia capillaris Kellogg	*Johnson 93104*, WS	*matK*	L34181	6
Gilia capitata Sims	*Johnson 92015*, WS	*matK*	L34182	6
Gilia caruifolia Abrams	*Johnson 93096*, WS	*matK*	L34183	6
Gilia filiformis Gray	*Johnson 93015*, WS	*matK*	L34185	6
Gilia hutchinsifolia Rydb.	*Johnson 93069*, WS	*matK*	L34186	6
Gilia inconspicua (Smith) Sweet	*R. Johnson 149*, WS	*matK*	L34187	6
Gilia leptalea (Gray) Greene	*Patterson s.n.*, WS	*matK*	L34195	6
Gilia leptantha Parish	*Schultz 52503*, WS	*matK*	L34197	6
Gilia leptomeria Gray	*Johnson 93008*, WS	*matK*	L34196	6
Gilia ochroleuca Jones	*Johnson 92022*, WS	*matK*	L34189	6
Gilia scopulorum Jones	*R. Johnson 304*, WS	*matK*	L34190	6
Gilia sinuata Benth.	*Johnson 92004*, WS	*matK*	L34198	6
Gilia splendens Dougl. ex Lindl.	*Johnson 92093*, WS	*matK*	L34191	6
Gilia stellata Heller	*Johnson 93059*, WS	*matK*	L34199	6
Gilia tenerrima Gray	*Johnson 93103*, WS	*matK*	L34192	6
Ipomopsis aggregata (Pursh) Grant	*Johnson 92100*, WS	*matK*	L34193	6
Ipomopsis congesta (Hook.) Grant	*R. Johnson 166*, WS	*matK*	L34200	6
Ipomopsis polycladon (Torrey) Grant	*Johnson 93068*, WS	*matK*	L34194	6
Langloisia setosissima (Torr. & Gray) Greene	*Johnson 93074*, WS	*matK*	L34201	6
Navarretia intertexta (Benth.) Hook.	*Glazner 9349*, WS	*matK*	L34202	6
Phlox gracilis Greene	*Johnson 92046*, WS	*matK*	L34203	6
Phlox hoodii Richardson	*Johnson 92001*, WS	*matK*	L34205	6
Polemonium californicum Eastw.	*Johnson 93089*, WS	*matK*	L34204	6
SAXIFRAGACEAE				
Astilbe japonica × *Chinesensis*	*Johnson s.n.*, WS	*matK*	L34114	2, 4
Astilbe taquetii (Léveillé) Koidzumi	Morgan & Soltis, 1993	*rbcL*	L11173	4, 5
	Soltis et al., 1990	r. s.	—	4, 5
Astilboides tabularis (Hemsl.) Engl.	Univ. Oslo Bot. Gard., Norway, O	*matK*	L34115	2, 3, 4, 5/3
	Soltis et al., 1993	*rbcL*	U06207	3, 4, 5/3
	Soltis et al., 1993	r. s.	—	3, 4, 5

APPENDIX 1. Continued.

Family Species	Voucher/citation	Data type	GenBank	Figure/Table
Bensoniella oregona (Abrams & Bacig.) Morton	*Soltis & Soltis, s.n.*, WS	*mat*K	L34112	2, 3, 4, 5/3
	Soltis & Soltis, s.n., WS	*rbc*L	L34072	3, 4, 5/3
	Soltis et al., 1991	r. s.	—	3, 4, 5
Bergenia cordifolia (Haw.) A. Br.	Komarov Bot. Inst., Leningrad, Russia, LE	*mat*K	L34116	2, 3, 4, 5/3
	Soltis et al., 1993	*rbc*L	U06208	3, 4, 5/3
	Soltis et al., 1993	r. s.	—	3, 4, 5
Bolandra oregana Wats.	*Grable 11587*, WS	*mat*K	L34117	2, 3, 4, 5/3
	Soltis et al., 1993	*rbc*L	U06209	3, 4, 5/3
	Soltis et al., 1993	r. s.	—	3, 4, 5
Boykinia occidentalis Torr. & Gray	*Grable 11636*, WS	*mat*K	L34118	2, 5
	Soltis et al., 1993	r. s.	—	5
Boykinia rotundifolia Parry	*Gornal 0101*, UBC	*mat*K	L34119	2, 3, 4, 5/3
	Morgan & Soltis, 1993	*rbc*L	L11175	3, 4, 5/3
	Soltis et al., 1993	r. s.	—	3, 4, 5
Chrysosplenium americanum Schw.	Soltis et al., 1993	r. s.	—	4
Chrysosplenium iowense Rydb.	*Wendel s.n.*, ISC	*mat*K	L34120	2, 4, 5
	Johnson & Soltis, 1994	*rbc*L	L19935	4, 5
Chrysosplenium tetrandrum (Lund) Fries	*Straly 6205*, UBC	*mat*K	L34121	2
Conimitella williamsii (Eaton) Rydb.	*Soltis & Soltis 1608*, WS	*mat*K	L34122	2, 5
	Soltis et al., 1991	r. s.	—	5
Darmera peltata (Torr.) Voss	*Soltis & Soltis 2083*, WS	*mat*K	L34123	2, 3, 4, 5/3
	Morgan & Soltis, 1993	*rbc*L	L11180	3, 4, 5/3
	Soltis et al., 1993	r. s.	—	3, 4, 5
Elmera racemosa (Wats.) Rydb.	*Soltis & Soltis 2234*, WS	*mat*K	L34124	2, 3, 4, 5/3
	Soltis et al., 1993	*rbc*L	U06210	3, 4, 5/3
	Soltis et al., 1991	r. s.	—	3, 4, 5
Heuchera hirsutissima Rosend., Butt & Lak.	*Wallace s.n.*, RSA	*mat*K	L34125	2, 5
	Soltis et al., 1991	r. s.	—	5
Heuchera micrantha Dougl.	*Soltis & Soltis 1949*, WS	*mat*K	L34126	2, 3, 4, 5/3
	Soltis et al., 1990	*rbc*L	L01925	3, 4, 5/3
	Soltis et al., 1991	r. s.	—	3, 4, 5
Heuchera rubescens Torr.	Soltis et al., 1991	*mat*K	L34127	2, 5
	Soltis et al., 1991	r. s.	—	5
Jepsonia parryi (Torr.) Small	*Rieseberg* 1110, RSA	*mat*K	L34128	2, 3, 4, 5/3
	Soltis et al., 1993	*rbc*L	U06211	3, 4, 5/3
	Soltis et al., 1993	r. s.	—	3, 4, 5
Leptarrhena pyrolifolia (D. Don) R. Br.	*Soltis & Soltis 2237*, WS	*mat*K	L34129	2, 3, 4, 5/3
	Morgan & Soltis, 1993	*rbc*L	L11191	3, 4, 5/3
	Soltis et al., 1993	r. s.	—	3, 4, 5
Lithophragma affine Gray	*Pellmyr & Thompson s.n.*, WS	*mat*K	L34130	2, 5
	Soltis et al., 1991	r. s.	—	5
Mitella caulescens Nutt.	*Soltis & Soltis 1881*, WS	*mat*K	L34131	2, 5
	Soltis et al., 1991	r. s.	—	5
Mitella diphylla L.	*Soltis & Soltis 1857*, WS	*mat*K	L34132	2, 5
	Soltis et al., 1991	r. s.	—	5
Mitella diversifolia Greene	*Soltis & Soltis 1910*, WS	*mat*K	L34133	2, 5
	Soltis et al., 1991	r. s.	—	5
Mitella nuda L.	*Johnson & Brunsfeld 1908*, WS	*mat*K	L34134	2, 5
	Soltis et al., 1991	r. s.	—	5
Mitella pentandra Hook.	*Grable 11432*, WS	*mat*K	L34135	2, 5
	Soltis et al., 1991	r. s.	—	5
Mitella stauropetala Piper	*Soltis & Soltis 1856*, WS	*mat*K	L34136	2, 5
	Soltis et al., 1991	r. s.	—	5

APPENDIX 1. Continued.

Family Species	Voucher/citation	Data type	GenBank	Figure/Table
Mukdenia rosii (Oliver) Koidzumi	*Soltis s.n.*, WS	*mat*K	L34137	2, 3, 4, 5/3
	Soltis et al., 1993	*rbc*L	U06212	3, 4, 5/3
	Soltis et al., 1993	r. s.	—	3, 4, 5
Peltoboykinia tellimoides (Maxim.) Hara	*Soltis s.n.*, WS	*mat*K	L34138	2, 3, 4, 5/3
	Soltis et al., 1993	*rbc*L	U06213	3, 4, 5/3
	Soltis et al., 1993	r. s.	—	3, 4, 5
Rodgersia pinnata Franch.	*Soltis s.n.*, WS	*mat*K	L34139	2, 3, 4, 5/3
	Soltis et al., 1993	*rbc*L	U06214	3, 4, 5/3
	Soltis et al., 1993	r. s.	—	3, 4, 5
Saxifraga cernua L.	Soltis et al., 1993	*mat*K	L34140	2, 5/3
	Soltis et al., 1993	*rbc*L	U06215	5/3
Saxifraga ferruginea Grah.	Soltis et al., 1993	*mat*K	L34141	2, 5
	Soltis et al., 1993	r. s.	—	5
Saxifraga integrifolia Hook.	*Soltis & Soltis 2253*, WS	*mat*K	L20131	2, 5/3
	Morgan & Soltis, 1993	*rbc*L	L01953	5/3
Saxifraga mertensiana Bong.	*Grable 11586*, WS	*mat*K	L34142	2, 3, 4, 5/3
	Soltis et al., 1993	*rbc*L	U06216	3, 4, 5/3
	Soltis et al., 1993	r. s.	—	3, 4, 5
Saxifraga oppositifolia L.	Soltis et al., 1993	*mat*K	L34143	2, 5/3
	Soltis et al., 1993	*rbc*L	U06217	5/3
Saxifraga punctata L. [= *S. nelsoniana* D. Don]	Soltis et al., 1993	*mat*K	L34144	2, 3, 4, 5/3
	Soltis et al., 1993	*rbc*L	U06218	3, 4, 5/3
	Soltis et al., 1993	r. s.	—	3, 4, 5
Suksdorfia ranunculifolia (Hook.) Engl.	*Soltis & Soltis 2308*, WS	*mat*K	L34145	2, 5
	Soltis et al., 1993	r. s.	—	5
Suksdorfia violacea A. Gray	*Soltis & Soltis 2309*, WS	*mat*K	L34146	2, 5
	Soltis et al., 1993	r. s.	—	5
Sullivantia oregana Wats.	Soltis et al., 1993	*mat*K	L34113	2, 3, 4, 5/3
	Soltis et al., 1993	*rbc*L	U06219	3, 4, 5/3
	Soltis et al., 1993	r. s.	—	3, 4, 5
Sullivantia sullivantii (Torr. & Gray) Britt.	*Quackenbush s.n.*, WS	*mat*K	L20130	2
Tanakaea radicans Franch.	Univ. Brit. Columbia Bot. Gard., UBC	*mat*K	L34147	2, 3, 4, 5/3
	Soltis et al., 1993	*rbc*L	U06220	3, 4, 5/3
	Soltis et al., 1993	r. s.	—	3, 4, 5
Telesonix heucheriformis Rydb.	*Wolf 151*, WS	*mat*K	L34148	2, 3, 4, 5/3
	Soltis et al., 1993	*rbc*L	U06221	3, 4, 5/3
	Soltis et al., 1993	r. s.	—	3, 4, 5
Tellima grandiflora (Pursh) Dougl. [= "northern" type]	*Soltis & Soltis 2113*, WS	*mat*K	L34149	2, 5
	Soltis et al., 1993	r. s.	—	5
Tellima grandiflora (Pursh) Dougl. [= "southern" type]	*Soltis & Soltis 2119*, WS	*mat*K	L34150	2, 3, 4, 5/3
	Soltis et al., 1993	*rbc*L	U06222	3, 4, 5/3
	Soltis et al., 1991	r. s.	—	3, 4, 5
Tiarella trifoliata L.	*Ness 533*, WS	*mat*K	L34151	2, 5
	Soltis et al., 1991	r. s.	—	5
Tolmiea menziesii (Pursh) Torr. & Gray	*Soltis & Soltis 1903*, WS	*mat*K	L34152	2, 3, 4, 5/3
	Soltis et al., 1993	*rbc*L	U06223	3, 4, 5/3
	Soltis et al., 1991	r. s.	—	3, 4, 5

APPENDIX 2. Insertion/deletion events (indels) observed in *mat*K sequence matrices following alignment. Indels are labeled alphabetically (A–D for *Gilia* and A–J for Saxifragaceae); these labels correspond to those mapped on the most parsimonious trees (Fig. 6 for *Gilia* and Fig. 2 for Saxifragaceae). Dashes represent missing bases associated with indels. Dots in the sequence below the reference taxon (*Gilia stellata* for *Gilia* and *Sullivantia oregana* for Saxifragaceae) indicate that the same nucleotide present in the reference taxon is also present in the species containing the indel (note that all species in which an indel occurs are listed, although the sequence is given for only one of these species when more than one species possesses the same indel). P/A represents the number of species with nucleotides involved in indels present and absent, respectively. RN, the reference nucleotide, is the position of the last sequential nucleotide preceding the indel in unaligned sequences of the reference taxon for each data set.

Data set Indel	P/A	Representative sp.	RN	Sequence region
Gilia				
A	30A	*Gilia stellata*	270	CAAAATCACATT TTTGGGCACAAC ------ACGAAT TTGTATTATCAA
	1P	*Polemonium californ.*		.... AACAAC..... .....C....
B	30P	*Gilia stellata*	378	GAGTTAGTCAAA TCTCATAATTTA CGATCAATTCAT TCAATATTTCCT
	1A	*Bonplandia geminifl.*		A.A..... ...--------- ....
C	26P	*Gilia stellata*	573	CACGAATATCGT AATTGGAATAAT ATTATTACTACA AAAAAATCTAGT
	5A	*Gillia leptalea*		.....G.... .....--- ---..... C
		Gilia capillaris		
		Allophyllum gilioides		
		Collomia linearis		
		Navarretia intertexta		
D	30A	*Gilia stellata*	798	GAAGTATTTATT AAAGATCTTCAA ---GATATTCTA GGTTTGTTCAAG
	1P	*Cantua pyrifolia*		.....C. ..... TCGTCG..... ..G.....A
Saxifragaceae				
A	44P	*Sullivantia oregana*	105	TATGCACTTGCT CATGATCATGTT TTAAATAGATCC ATTTTGTTGGAT
	1A	*Tetracarpaea tasman.*		..... T.....--- ------.....A TG.....A
B	43A	*Sullivantia oregana*	141	ATTTTGTTGGAT AATTTTGGT--- TATGACAATAAA TCCAGTTCAATA
	2P	*Chrysosplenium iowe.*		.C...T...A.. ...A.....GGT ...T..... .....C..
		Chrysosplenium teta.		
C	44P	*Sullivantia oregana*	237	ATTTCTGCTAAT TATTCGAACAAA AATAAATTTTTG GGGCATAACAAG
	1A	*Suksdorfia ranuncul.*		..... G..... ---------... .....
D	42A	*Sullivantia oregana*	276	AAGAATTTCGAT TCTCAA------ ATGATATCAGAG GGATTTGCAGTT
	3P	*Saxifraga integrifo.*		.....A... .A...AATAAA C
		Saxifraga punctata		
		Saxifraga ferrugine.		
E	42A	*Sullivantia oregana*	354	TCTTCCTTAGAA AGGAAAGAA--- ATAGTAAAATCT CATAATTTACGA
	3P	*Saxifraga cernua*		.....G..... .....GAA ....G..... .....
		Saxifraga oppositif.		
		Astilbe jap. × chin.		
F	44P	*Sullivantia oregana*	366	GAAAGGAAAGAA ATAGTAAAATCT ------CATAAT TTACGATCAATT
	1A	*Saxifraga mertensia.*		.....A..... .....C.....G CAAAAT..A... .....C...
G	40P	*Sullivantia oregana*	366	GAAAGGAAAGAA ATAGTAAAATCT CATAATTTACGA TCAATTCATTCA
	5A	*Heuchera micrantha*		....A..... T..... ------..... .....
		Conimitella willia.		
		Mitella diversifol.		
		Mitella stauropeta.		
		Tellima grandif. S.		
H	43A	*Sullivantia oregana*	549	TTTCTCTATGAG TAT------CAG AGTTGGAATAGT CTTATTACCCCA
	2P	*Chrysosplenium iowe.*		.....T..... ...TATATTT.C .A..TA..C.A. .....
		Chrysosplenium teta.		
I	44P	*Sullivantia oregana*	573	AGTTGGAATAGT CTTATTACCCCA ACTCCAAAGAAA TCCATTTCCATT
	1A	*Tetracarpaea tasman.*		.A..... .....T--- ---..... .....T...
J	43P	*Sullivantia oregana*	591	CCAACTCCAAAG AAATCCATTTCC ATTGTTTCACAA AGGAATCAAAGA
	2A	*Leptarrhena pyrolif.*		.....T ......------ .....A.. .....
		Tanakaea radicans		